STATISTICS IN PSYCHOLOGY AND THE LIMITATIONS OF THE TEST METHOD

BY E. G. CHAMBERS, *From the Cambridge Psychological Laboratory*

I. *Introduction* (p. 189). II. *The concept of quantity as applied to psychological data* (p. 189). III. *Use of statistical methods* (p. 191). IV. *Test scores and the weighting of scores* (p. 192). V. *The meaning of test scores* (p. 194). VI. *Common factors* (p. 194). VII. *Measures of practical ability* (p. 195). VIII. *Norms of intelligence in industry* (p. 197). IX. *The future* (p. 199).

I. INTRODUCTION

The young science of experimental psychology is tending more and more in these days to become a precocious child. Increasingly it is attempting to imitate its elder brethren, the so-called exact sciences, by concerning itself with quantities and mathematics; and it is time that psychologists paused and considered whither they are tending and what are the assumptions on which so much of their work is based. For statistics should be the servant of science, and there are disquieting signs that in psychology the servant is becoming the master. Too frequently, it seems, mathematical psychologists build elegant and dizzy numerical edifices, forgetting in their architectural zeal the flimsy foundations upon which their fabrics stand.

When the present situation is examined two questions of fundamental importance emerge immediately: How far can the concept of *quantity* be applied to psychological data? And how far, if there is such a thing as quantity in psychology, is it amenable to treatment by ordinary arithmetic methods?

II. THE CONCEPT OF QUANTITY AS APPLIED TO PSYCHOLOGICAL DATA

Let us examine the first question, that of *quantity* in psychological data. The writer feels that the consideration of this question is too frequently shirked and that too often substitutes which may be treated mathematically are uncritically accepted in place of genuine psychological data. Suppose we take as a psychological datum the statement "Abelard loved Heloïse'. That is a psychological fact at which no psychologist or statistician or psychiatrist would cavil. If we go further to ask "How much did Abelard love Heloïse?" we at once get into difficulties and have recourse to such vague statements as that he loved her 'deeply' or 'to distraction'. No statistician would be bold enough to assert that his love for her was removed from the average by nearly four times the standard deviation—an event which would occur by chance only once in twenty thousand times! If, moreover, we also ask "Did Abelard love Heloïse as much as Dante loved Beatrice or as David loved Jonathan?" we involve ourselves in a good deal of trouble and are forced to state that the *quality* of the love was different in each case and that any attempt to assess and compare *quantities* of love is absurd.

Now here we have chosen an extreme case, and admittedly no statistical psychologist has yet dared to assess the amount of love shown by a hundred husbands for their wives on a five- or ten-point scale. But persons have been asked to give marks to their emotional moods on a scale ranging from zero to ten and so to arrange the marks that they form a 'normal' distribution! Is this so very far removed from the absurdity of treating love as a commodity which may be weighed and made into parcels and tied up with pink

ribbon? The only way in which an individual's marking of his own moods can be regarded as a psychological datum is to state that a certain individual with an unknown capacity for introspection, with an unknown degree of understanding of the words used to categorize the moods, and subject to highly artificial restrictions in his use of marks, on one certain occasion produced a particular set of marks. Some slight use might be made of such a marking system if the individual were asked to re-mark himself on a number of occasions, on a Monday morning, during an air-raid, after a bottle of wine, etc. Comparison of the sets of marks so produced might conceivably throw some light on the ebb and flow of emotional moods in that individual, or conversely they might give some indication of the reliability or unreliability of the individual in assessing his own emotional make-up; but to regard the marks as meaningful measures of quantity and to subject them to mathematical treatment on that assumption is, to say the least, an exceedingly doubtful procedure.

Such artificial attempts to obtain figures for mathematical jugglery are fortunately still not common, although they appear to be on the increase; but is the position so very different when we come to consider the widespread use of psychological tests? To define the situation more narrowly, let us consider a particular 'intelligence' test, X, which is given to a number of subjects. The first point to note is that although it is usually tacitly admitted that we do not know what test X is 'measuring', in practice the test is commonly referred to as an 'intelligence' test, which, from the psychological point of view, is begging the question, unless intelligence be defined as the ability to perform test X. If the latter were the case, then it follows that there are as many forms of intelligence as there are 'intelligence' tests (unless there is a perfect correlation between these tests) and this is highly improbable. What is much more likely to be the true state of affairs, in the writer's opinion, is that each individual possesses a complex of qualities enabling him to behave in a manner which is to some degree 'intelligent'. This complex of qualities is living, so that it may expand, contract or otherwise change, and since, like every other psychological quality, it does not and cannot function in isolation, it functions with varying degrees of imperfection on different occasions. Thus it is very difficult to see what *quantity* of intelligence can mean—certainly a score obtained on one 'intelligence' test on a single occasion cannot be taken as 'measuring' the amount of intelligence a person possesses, except in a very narrowly restricted sense.

No one will dispute, however, that some meaning may be attached to the statement that "A is more intelligent than B", in the sense that A manifests more intelligence than B in his behaviour, which may or may not include the performance of 'intelligence' tests. The use of the comparative 'more' suggests that some idea of quantity lies behind the statement but it can scarcely be quantity in the ordinary connotation of the term, since no one would say "A is exactly one and a quarter times as intelligent as B". Moreover, the statement is true only in a certain type of situation, for example, one demanding ability to reason syllogistically; or else as a sort of average, taking all types of general behaviour together. In certain types of situation A might be described as 'rather stupid', but he would not be so described as frequently as B over a wide range of situations. Thus A may manifest his intelligence in situations demanding verbal facility, mathematical insight, etc., but he may appear less intelligent when, say, an appreciation of spatial relationships is required. Once again, therefore, it seems as though it is quality of intelligence rather than quantity that differs from person to person; unless we could take as a measure the number of different situations in which an individual

showed intelligence. This last 'measure' is scarcely practical since it would be impossible to observe large numbers of people in a great variety of situations, each situation making equal demands on the 'intelligence' of individuals. Hence arises the practice of relying on a few test situations and taking the scores on these tests as a 'measure' of amount of intelligence.

This practice is attended by certain grave dangers. In the first place it leads to a strong tendency to regard psychological qualities in isolation; this is a retrograde step, a reversion to the days of 'faculty psychology'. An individual functions as an integrated whole and not as a sum of discrete qualities. He does not react to a certain situation merely by using a particular quality, say mechanical aptitude, in isolation: he reacts as a unified personality, and his success in reaction may very easily depend on his possession of other qualities not apparently demanded by the situation. In the psychological make-up of an individual the whole is strikingly more than the sum of the parts, and any attempt to portray a person's mentality by the sum of test scores or 'factors' can have only the sketchiest success. Secondly, in the absence of any direct measure of psychological qualities and in complete ignorance of the sort of units in which such qualities might be measured, test scores are seized upon and subjected to all kinds of mathematical treatment. Now in so far as test scores are sets of varying numbers they are legitimate material for statistical manipulation, provided always that they conform to the mathematical requirements of the statistical methods employed. This last point is too frequently overlooked or, even worse, the test scores are themselves arbitrarily changed so as superficially at least to meet mathematical requirements. Thus a group of subjects is given an 'intelligence' test consisting of a series of small items. An individual's score might be taken as the number of items correctly answered, and on this basis the subjects could be ranked in order of merit. Very often, however, the experimenter is not content to leave matters like this. Arguing that certain items are more difficult than others and so deserve more marks, he assigns a set of weighted marks to the different items, the basis of weighting sometimes being his own estimate of relative difficulty. Now this produces a different set of marks for the same subjects at the same test and may very easily rank the subjects in quite different order from that obtained by the simpler method of assigning one mark per item. Any statistical work based on these two sets of marks will yield different results, and how can the experimenter possibly tell which marking system approximates more nearly to the truth? That which yields higher correlation coefficients is, of course, usually more satisfactory from his point of view, but it is only *marks* which he is correlating, and the question as to whether these marks do in fact measure something stable and capable of linear measurement in the subject is still left open.

To sum up the foregoing paragraphs we may say that it is at least very doubtful whether the concept of measurable quantity may be applied at all to psychological qualities; it is certain that psychological qualities do not function in isolation; and it is exceedingly doubtful what is the significance of marks arbitrarily assigned to the results of a single testing on a particular occasion. It is on this very uncertain basis that the whole of the work of the factorial analysts rests.

III. USE OF STATISTICAL METHODS

So far little has been said about the actual use of statistical methods in psychological work. Broadly speaking, psychologists may be divided into four classes in their attitude towards statistics. The first class knows very little about mathematics and tends to

dislike and mistrust statistical work. The members of this class are not necessarily mere armchair psychologists: they may be very excellent experimentalists and realize that the behaviour of a single individual on a single occasion may provide real psychological data, although we may as yet be ignorant of the various factors entering into that individual's reactions. In the second class fall those who, having some mathematical ability, may yet be relatively ignorant of the theory and assumptions underlying statistical method. These tend to regard statistical work as a kind of magic which produces nicely proved results from possibly ill-designed experiments. Thirdly, there are those who are skilled mathematicians and like to exercise their skill in the psychological field. To these actual experiment tends to be a perhaps annoying necessity for the production of figures for their expert manipulation. The remainder of psychologists includes those who first ensure that the measure they are using has meaning and who design their experiments in such a way that they yield data which may legitimately be treated by the methods of statistics. These psychologists recognize statistical method for what it is—a tool with a two-fold use, first, that of describing in brief form the data obtained in experiment, and secondly, that of examining hypotheses made about those data. No estimate of the proportion of psychologists falling into these four classes is attempted.

There are certain considerations in the use of statistics which are not apparently generally realized, and a brief indication of a few of the more important of these may be of value.

(1) Statistics deal only with *numbers* and the results of statistical procedure remain numbers. The meaning and interpretation of these numbers, both the original data and the final results, are a matter for the science concerned and not for the mathematician.

(2) Statistics are essentially applied to numbers on a linear scale, and before statistical work can be started it is imperative to verify that any measure used is linear. Time scores, numbers of items answered, numbers of individuals in different classes, etc., are linear, but it does not follow by any means that all (or any) psychological qualities are capable of linear measurement. Proof of linearity must be obtained in cases of preferences, judgements, estimates of mood intensities, etc., before the question of quantity and its measurement can even arise.

(3) Each different statistical method is applicable only to certain types of data, and it is essential to ensure that the requisite conditions for use are satisfied before embarking upon a statistical calculation. For instance, the method of product-moment correlation requires that the two variables under examination shall be normally distributed and that their regressions on each other shall be linear.

(4) Direct proof of an hypothesis is impossible by statistical means; only disproof is possible. Tests of significance are based on known distributions of data and depend ultimately on probabilities. Thus we might find that the observed relationship between two variables would arise by chance only once in a hundred trials. This may be taken as a definite disproof of the null hypothesis that there is *no* relationship between the two variables and logically, therefore, there must be some relation between them. It does not prove the existence of that relationship directly, however, since it is possible, if unlikely, that the particular experiment concerned did yield the one chance result.

IV. Test scores and the weighting of scores

Certain points which have been dealt with very briefly so far merit further discussion. An important one of these is the problem of scoring and weighting the scores of tests. Certain scores, such as the times of performance of a standard task, are obvious and

relatively meaningful. The only drawback from the statistical point of view is that time scores are invariably distributed in a skewed form thus making them unsuitable, on a strict view, for some sorts of statistical treatment which assume a normal distribution, e.g. the calculation of correlation coefficients. However, the skew is not as a rule very pronounced, and statistical methods may generally be applied to time scores with the exercise of a little caution in interpreting the results obtained. Scores representing the absolute number of tasks of equal difficulty performed in a fixed time (for example, the number of pegs inserted in a pegboard in three minutes) also usually have a real meaning and are amenable to statistical treatment.

The problem arises most urgently when we come to paper and pencil tests which comprise a series of items of different difficulty. Were all the items of equal difficulty, the number correctly answered in a fixed time would be an adequate score of test performance, but it is usually impossible to ensure this in practice, except in the case of the simplest tests. Much more often the items differ in difficulty and are frequently graded deliberately in order of difficulty. What method of scoring has meaning in such a case? (We may leave out of consideration for the moment the type of frequency distribution curves of scores yielded by the different possible methods of allotting marks, since this question is of importance only when proceeding to statistical work and does not enter into a consideration of the meaning of scores. The idea of setting up a marking system with the sole aim of producing a normal distribution of scores cannot be too strongly deprecated: it provides a glaring example of the misplacement of emphasis from the securing of meaningful data to mathematical convenience.) Now to give one mark to each item correctly answered has the merit of simplicity, but this method seems unfair since certain items may demand more of the ability being 'measured' than others and so their correct answering should receive adequate recognition. The allocation of different marks according to the test deviser's idea of relative difficulty is not easy to justify: such a system must be dependent in any case on the intelligence of the tester himself and the types of question he finds difficult.

Perhaps the most usual method of score weighting is to base the marks for each question on the frequency with which that question is correctly answered by a large group of subjects, a question which is correctly answered only half as many times as another being regarded as twice as difficult and accordingly given twice as many marks. As usually carried out this method involves a good deal of approximation. Questions towards the end of the test are much less often attempted owing to shortage of time, and on the whole only the more intelligent subjects will get far enough to attempt them; since such subjects are more intelligent they will tend to answer these later questions correctly and so these items will appear easier than they would had they occurred earlier in the test and been attempted also by less intelligent subjects. Moreover, subjects may spend a fair time on particular items and pass on leaving no indication of their attempts—another factor which is usually left out of account in the method of calculating the proportion of correct answers to attempts. Also, such proportions will vary a good deal so that marks should strictly be fractional in many cases; practical convenience, however, generally leads to the allocation of integral scores. Strictly speaking, the correct method of weighting scores by the proportion of right answers to attempts is to give every subject the same time to attempt each separate item. Number of correct answers alone would then provide an adequate basis for assessing relative difficulty. This method is rarely, if ever, adopted in practice. (Even this leaves out of account the effects of the order in which the items are presented, of practice on the earlier items and fatigue on the latter ones.)

V. The meaning of test scores

Suppose that a reasonable system of weighting item scores has been adopted and that, to take a simple case, certain items receive one mark, others receive two and the remainder three; then a subject who correctly answers three of the most difficult items will gain exactly the same number of marks as a subject who correctly answers nine of the easiest items. From a statistical point of view these two subjects are then exactly equal, and in any statistical treatment, correlation, factorial analysis or what you will, they remain exactly equal. Can they be said to be equal in any sense, however, from the psychological point of view? Do three problems involving syllogistic reasoning make exactly the same demands on 'intelligence' as nine problems involving simple analogies? Does it require exactly the same amount of the same sort of intelligence to supply two words opposite in meaning to two given words as to solve one problem involving an anagram or the rearrangement of words in a sentence? The treatment of subjects with equal marks as equal in intelligence assumes an affirmative answer to these questions, but surely the assumption cannot be seriously maintained? It omits, for example, differences in interest between subjects, the type of interest which will make one subject happier to solve or even attempt one difficult problem that appeals to him than four simple problems which have no interest for him. Once again it must be stated categorically that an individual reacts to a situation (which may be doing an intelligence test) as an integrated whole, and to regard the score on such a test as 'measuring' one quality in isolation is absurd and misleading.

As a further example of the fallacy of taking equal scores in a test to indicate possession of an equal amount of some measurable quality let us take the case of a so-called test of mechanical aptitude. Equal scores may be obtained by individuals in such a test in several different ways. One may really have a mechanical 'sense' and be able to solve the problems by virtue of this capacity; another may solve an equal number of problems by the exercise of logical reasoning on unfamiliar material; a third may have a capacity for appreciating spatial relationships and be able to solve mechanical problems by arguments based on such relationships. Here again these individuals cannot be regarded as equal in any sense except from the sole point of view that they have gained equal marks on a particular test; yet statistical work would assume that they possessed equal quantities of some capacity called (begging the question) 'mechanical aptitude'.

VI. Common factors

Following this the question arises: Does not the use of a battery of tests having a 'common factor' throw more light on what we are attempting to measure? In answering this we have first to consider the meaning of 'common factor', and it may be stated definitely at the outset that establishing by methods of correlation the existence of a 'common factor' between a number of tests signifies nothing more nor less than that there is a *mathematical* factor common to the test *scores*. It cannot by itself add one iota to our knowledge, or ignorance, of what each individual test score signifies. Were each test score a quantitative measure of some isolated ability which, and which alone, was functioning at the time of testing, then a correlation between two test scores would in fact indicate the degree to which individual subjects tended to possess related quantities of two isolated abilities. But this provision, as we have seen, cannot be allowed for one moment so that the multiplication of tests and the whole elaborate paraphernalia of

factorial analysis can still add nothing to our theoretical knowledge of human abilities. (The practical value of the test method will be considered later.)

Suppose we go yet a stage further and postulate not only a set of intercorrelating tests having a common factor but also a numerical measure of some practical ability which correlates with the test scores and has the same common factor; would not this throw considerable light on the nature of the common factor and the abilities being 'measured' by the tests? The answer is once again in the negative, since we do not know what it is in the individuals concerned which causes their varying degrees of success in the practical work. Once again the individuals react to the practical situation as integrated wholes, and the fact that two individuals are graded as equal in practical ability does not necessarily mean that they have the same amount of some single ability or some cluster of abilities in stated proportions. There is, for example, considerable evidence to show that the qualities required for the successful performance of a skilled task by an expert are not the same as those required by someone learning to perform that same task, so that we cannot say that a skilled task demands the exercise of certain abilities without defining the stage of proficiency of the workers; and even then we do not actually know what are the qualities required at the various stages of learning and proficiency.

From all this the writer draws the deduction that theoretical knowledge of the psychology of the individual cannot be advanced by the use of tests and the statistical manipulation of test scores. That this is true from the statistical standpoint is fairly obvious. Statistics essentially relate to groups and not to individuals. For example, if we know the correlation between two variables, A and B, and their regressions on each other, then a knowledge of the amount of A possessed by an individual will enable us to prophesy how much B he will possess. But our prophecy will not necessarily be correct: it will have only a certain chance of correctness within certain limits of variability. If the correlation between A and B is very high we might, for instance, be right *on the average* in 80% of our prognostications, but we should still be wrong in the remainder of the cases, and further there would be no indication from the statistics which were the individuals about whom we were wrong. In practice, if A were a selective test score and B a measure of some practical ability, a relationship of this degree would be very useful, but we could not point to a single individual and confidently make the definite assertion that since his score in A was high therefore he would excel in the ability measured by B. Thus psychological tests may usefully be applied from a pragmatic point of view: given to groups of subjects it does not in the least matter what the test scores correspond to in the individual subjects—as long as the scores consistently correlate with some other measure a useful selective purpose may be served. But still no light is thrown on what the tests are measuring and the assumption has still to be made that equal scores correspond to something equal in the individuals.

VII. MEASURES OF PRACTICAL ABILITY

It has been implicitly assumed in this argument that some numerical measure of practical ability is available. In practice a valid measure of this sort is usually exceedingly difficult to obtain. The first intelligence tests were 'validated' by giving them to students or to school children and correlating the test scores with the marks obtained by the same subjects in examinations. In this connexion two questions arise: (1) Do school examinations measure anything more than the ability to succeed at school examinations? (2) Is

there any real evidence that people who do well in school examinations subsequently succeed in non-scholastic vocations?

The consideration of the former question again raises a point previously mentioned in regard to test performance, namely, that scholars may gain equal marks in examinations for very different reasons. One may have a genuine *flair* for the subject of the examination; another may through interest and studious application succeed in being able to answer questions of the sort asked; a third, who needs to pass the examination for economic reasons, may get himself 'crammed' sufficiently to deal adequately with the paper, and so on. No one would maintain, however, that these different students are in any way equally equipped to face their future as the world's workers.

Concerning the second question, the writer knows of no evidence to prove that examination success is strikingly related to success in non-scholastic vocations. (Since entry into the scholastic vocations is only possible through examination success, the question does not arise in their case.) Indeed, strong support may be given to the view that entry into the professions and better paid occupations is dependent more on social factors than on scholastic attainments, although these may be by no means independent. Different occupations obviously differ in their demands on intelligence, but the difference lies in the type of intelligence demanded rather than the quantity, so that it would not be possible to rank occupations on a linear scale according to the *amount* of intelligence they require. For instance, an accountant, a solicitor, a scientific research worker and a motor mechanic each requires considerable intelligence to succeed in his own vocation, but it is a different sort of intelligence in each case, even if certain fundamental laws of ratiocination have to be obeyed by all, and the question of relative quantities of intelligence does not enter. It is partly in recognition of this view that selective tests for different occupations have been developed.

Now much work on selective tests of a psychological nature has been done, but it is very difficult to obtain concrete evidence of their value. The reason for this in a great many instances is the extreme difficulty of obtaining an adequate criterion of success or ability in the occupation itself. In certain instances, output, in the sense of number of objects produced in a fixed period, might be used as a measure, but the issue is often confused by incentives and rates of pay and the economic demands of the individual workers. Output by itself is often of little value as a measure of ability, since the maximum output allowed in many cases is frequently well within the reach of an inferior workman. Bricklaying gives a striking example of this factor. In other occupations, where payment is made on a time basis, output may be no indication whatever of ability, and in still others it seems to be impossible to obtain a criterion of ability at all, except as a matter of opinion. By what yard-stick shall we measure ability to drive a motor car, for example, or, more topically, ability to pilot an aeroplane? It is easy to recognize the failures—the churchyards are full of them!—but how can we discriminate between grades of efficiency amongst those who are at least reasonably expert?

This continually recurring difficulty too frequently leads to the adoption of criteria of very doubtful significance, and it is obvious that unless the criteria have real meaning, the value of tests which correlate with them is by no means proved. It is possible that the whole approach to the question of vocational selection is unsound, and that what should be aimed at is not a correlation throughout of test and ability but the fixing of a standard of personal ability (possibly determinable by tests) below which individuals are regarded as useless for the occupations in question. That real variations in rifle marksman-

ship exist is proved annually at Bisley, but the army undertakes to train every man enlisted in the infantry to shoot; the Service does, however, usually refuse to enlist a man with one arm or with gross optical defects. From the results of the haphazard methods of engaging workers which obtained so largely in the past one may legitimately deduce that the majority of mankind could, with adequate training, become reasonably efficient in the bulk of human occupations. It would seem, therefore, that the aim of selection should be the elimination of the incapable rather than the differentiation of everyone, and that the chief emphasis should be thrown on training. These remarks apply mainly to selection by *ability*: further selection of the capable on temperamental and social grounds would still be possible but fortunately we have not yet been called upon to 'measure' these two variables on a numerical basis.

VIII. NORMS OF INTELLIGENCE IN INDUSTRY

Finally, we come to a question of vital importance to the post-war world: Can the psychologist help in the reorganization of industry by establishing intelligence norms for different occupations? The assumption of an affirmative answer to this question has already been made in certain of the Services, and some psychologists think that the principle may be extended to embrace the whole of industry. There are, however, certain obstacles in the way of realizing such a possibility and some grave dangers attending the use of such 'norms' without very careful examination of their meaning and value.

In examining this question let us first put it in a more concrete form and ask: Is it possible by the use of a single 'intelligence' test to discover the average or minimum test score necessary to be obtained by the workers in each separate industrial process? Those who give a positive answer to this question envisage a list of occupations ranked in order of the amount of 'intelligence' demanded by each, so that persons scoring a mark X on the test would possess the necessary amount of intelligence for success in occupations A, B, C, ..., N, but insufficient for success in P, Q, R, ..., Z. Enough has been said already to indicate the writer's belief that it is *type* and not *quantity* of intelligence which differentiates occupations, so no more will be added here on that point. Instead, let us examine the methods by which the establishment of 'intelligence norms' for different industrial processes might be attempted.

There are two possible methods of attempting to establish the standards of intelligence required for success in different occupations:

(1) The first method would be to test a representative sample of workers at present in each occupation. This would give us the average and range of test scores for each separate occupation, but in order to use these data for establishing norms a number of assumptions would have to be made. It would have to be assumed:

(*a*) that the workers tested were in fact in the right jobs;

(*b*) that the average intelligence level manifested by the workers in each occupation is actually necessary for that occupation;

(*c*) that workers in a particular occupation are really capable of doing it, and that for real efficiency it would be unnecessary to obtain workers of a higher standard (if such were available);

(*d*) that the intelligence test score of the least intelligent worker in any job represents the minimum intelligence level for that job; and

(*e*) that workers in a particular occupation are not capable of succeeding in another

occupation which appears to require a higher standard of intelligence, as judged by the test performance of the workers at present in that other occupation.

Now these assumptions are by no means necessarily justifiable. For example, we have no evidence to show that workers are actually in the occupations for which they are best suited. Highly intelligent individuals tend to select their own occupations amongst those which are interesting, remunerative or carrying social prestige. It does not follow that the jobs which they select really demand the degree of intelligence which these individuals would manifest on testing. Moreover, the assumption that workers are in the highest occupation on the 'intelligence scale' that can be attained by people of their level of intelligence leaves entirely out of account the geographical, social and temperamental factors which enter into seeking employment. There is no proof either that a number of occupations could not be pursued more efficiently by workers with a higher average intelligence than those at present in those occupations, assuming that these more intelligent people were available.

On the whole, therefore, this method of attempting to establish intelligence norms for industry seems unlikely to succeed or to give much useful information if it were employed.

(2) The second method would be to test all new entrants into each occupation and then follow their subsequent careers of success or failure in those jobs. This method might in course of time indicate the minimum test score necessary for success in each occupation—'success' here meaning the avoidance of discharge through inefficiency. It would, however, leave out of account the fact that different social groups, with very different educational facilities, tend to apply for particular jobs for geographical and other reasons. Unfair comparisons between occupations would still obtain, therefore, since relatively unintelligent workers might not even apply for jobs which they could in fact do perfectly well, these jobs appearing as a result to demand a higher standard of intelligence than is really the case. The reverse may also be true; certain positions which are filled by persons with social rather than intellectual gifts might be more efficiently occupied by workers selected for their intelligence rather than by the colour of their ties.

A further point arises in connexion with the attempt to rank industrial occupations on the basis of intelligence. It is quite possible that the special aptitudes and skills demanded by different industrial processes outweigh in importance the degree of intelligence required. Persons who would rate as subnormal or even stupid on an intelligence test may become efficient bus drivers, shoemakers, typists, etc., by virtue of the acquisition of some skill. Although country villages may conceal "mute, inglorious Miltons", very few of the old craftsmen, with the exception of people like Benvenuto Cellini, made a name for themselves intellectually. In so far as this is the case, the value of any selection on the basis of intelligence alone is likely to be strictly limited and manifest only in cases where intelligence is the most important requirement of an occupation.

Even if it were possible to establish a minimum standard of intelligence required by each occupation, it still would not follow that differences in intelligence amongst those above that standard correspond to differences in ability in the occupation. It would seem far more likely that where a certain standard of intelligence obtains, relative success in an occupation depends more upon special aptitudes and temperamental factors than upon differences in intelligence, this naturally being dependent on the type of occupation. There is some evidence in favour of this view, but the attempt to substantiate it completely would throw us once again against the difficulty of obtaining a criterion of relative ability or success in industrial processes.

IX. THE FUTURE

In this paper the writer has attempted to examine some of the current methods of experimental and practical psychology, to point out their underlying assumptions and their dangers and to indicate what, in his opinion, is the limit of their usefulness. He views with considerable alarm departures from the strict scientific method of first collecting facts, secondly, making hypotheses about those facts, and thirdly, examining the truth of these hypotheses by experimental and statistical methods. In the stress of wartime short cuts and leaps in the dark may be necessary, but they do not constitute a desirable method of pursuing scientific truths. Too often the essential collection of facts is performed in a perfunctory manner, and the writer sincerely hopes that amongst the many plans for reconstruction in the post-war world one for the pursuit of psychological truths by scientific methods will not be forgotten.

Reasons have been adduced for the belief that some of the methods at present in use will add little or nothing to theoretical knowledge. Practically, however, the picture need not be so gloomy. If (it is rather a big 'if') industrialists can produce any criterion of ability or success which is consistent and which really does discriminate between individuals, and can further show that the differences measured by this criterion have economic or any other significance, then the psychologist can almost certainly devise tests the performance of which shall have a valuable selective relationship to the criterion. As stated previously, it would not at all matter what these tests really 'measured' as long as they always correlated with criteria which differentiated persons in some practically valuable manner. Further, without at all departing from the position which has been firmly held in this paper, that the individual functions as an integrated whole and not as a bundle of discrete qualities, we may say that in certain situations certain qualities will predominate in the individual's reactions, and for *practical* purposes tests may be regarded as 'measuring' these qualities. The writer believes, moreover, that even in the absence of a valid criterion of practical ability the psychologist could still improve the general standard of workers in an occupation by the use of selective tests, but he could never in this case prove the efficacy of his selection. Indirect evidence only would be available and the attempt to link such evidence of improvement with the tests would be exceedingly difficult.

In conclusion, the writer wishes gratefully to acknowledge the help and stimulus of discussions on these topics with Prof. F. C. Bartlett, who cannot, however, be held responsible for the opinions expressed herein.

(*Manuscript received* 10 *November* 1942)

PERSEVERATION AND FLUENCY

By BERNARD NOTCUTT, *From the Department of Psychology, Natal University College*

I. *Introduction and tests used* (p. 200). II. *Administration and scoring of tests:* (1) *Dark adaptation* (p. 201); (2) *After-image* (p. 201); (3) *Mirror drawing* (p. 201); (4) *Motor perseveration tests* (p. 202); (5) *Tongue twisters* (p. 203); (6) *Colour naming test* (p. 203); (7) *Free association test* (p. 203); (8) *Continuous association* (p. 203); (9) *Fluency tests* (p. 203); (10) *The 'W' factor* (p. 204); (11) *Surgency* (p. 204). III. *Results* (p. 204): (1) *Perseveration tests* (p. 204); (2) *Fluency tests* (p. 207); (3) *Relations to supposed factors of temperament and personality* (p. 207). IV. *Summary* (p. 208). *References* (p. 208).

I. Introduction and tests used

The following series of experiments was designed to study the interrelations of a number of supposed tests of perseveration and fluency, and their relations to temperamental and intellectual traits. Tests of perseveration were designed to measure four different aspects of this supposed factor. (*a*) *Sensory perseveration*: This is the 'secondary function' of Gross (1902), Wiersma (1906), and others. According to this theory individuals differ in the degree to which they are influenced by the effects of a sensory stimulus after the stimulus has ceased; and these individual differences correspond to differences in temperament. (*b*) *Motor perseveration* ('creative effort' type): This is measured by the ease with which subjects are able to break down a familiar habit, for instance, making S's backwards. (*c*) *Motor perseveration* ('alternation' type): This is measured by a comparison of the difficulty of performing two similar tasks independently, and in alternation. It is based on the hypothesis that individuals vary in the readiness with which they can shift from one activity to another. These two types of motor perseveration have usually been confused in the literature. Darroch (1938) has pointed out that most of the methods of scoring designed to measure (*c*), have really measured a mixture of (*b*) and (*c*), in variable and unknown proportions. We have designed a method of scoring which is free from this objection. (*d*) *Associative perseveration*: This is a tendency for ideas to occur involuntarily to the mind, once they have been aroused (Müller & Pilzecker, 1900).

Spearman (1926) sought to combine all these concepts in his 'Law of Inertia'. Spearman also sought to identify individual differences in perseveration with the antithesis of introversion-extraversion, and other similar typologies. We therefore gave tests of introversion, and of Webb's (1915) '*W*' factor, in order to study the relations of these to perseveration.

Measurements of 'fluency' were also obtained. Fluency is a measure of the rapidity of associative processes. Our tests were based on those of Studman (1935). Measurements of 'surgency' (Cattell, 1933) and intelligence were obtained, in order to study the relations of these to fluency.

The subjects were fifty educated adults, mostly part-time students at the Natal University College, Durban. The majority were primary school teachers. All the subjects took part in all the tests, most of which were administered individually, though a few were suitable for group work. The experiments were carried out by four post-graduate students of the Natal University College.[1]

[1] Those on sensory perseveration were done by Miss M. A. Cauvin; on motor perseveration by Miss M. B. Eckhoff; on associative perseveration by Miss H. Taitelbaum; on fluency by Miss C. B. Anthony. The author was responsible for the planning and direction of the experiments, and for part of the calculations.

The following measurements were obtained for each subject:

Quality to be measured	Experiment	Measure used
Sensory perseveration	Dark adaptation	Time
	Negative after-image	Time
Motor perseveration ('creative effort')	Mirror drawing (six-pointed star)	Distance covered accurately in unit time
	Two tests from Cattell (1935): '234' written normally, then with reverse stroke; and 'H' written normally, and horizontally, 'H'	Ratio of number written normally, to number written in new way, in unit time
Motor perseveration ('alternation')	Four tests from Cattell (1935); in addition to the two described above, we used 'ABCD' followed by 'abcd', and ± followed by ∞	Relation of number written in independent performance, to number written in alternating performance: formula is given below
	Tongue-twisters: three sentences involving difficulties in sequence of sounds	Number of trials required for three successful repetitions of each sentence
Associative perseveration	Stroop's (1935) colour naming test	Time
	Free associations to 50 words given twice	Frequency of repetition of same response to same stimulus
	Continuous association	Frequency of repetition of response
	Same	Mean length of trend
Fluency	Three tests modified from Studman (1935): controlled continuous association, interpretation of pictures, interpretation of ink-blots	Number of responses in unit time
	Free association	Median reaction time
	Continuous association	Number of words in unit time
Speed	Speed of writing '234', 'ABCD', 'abcd', 'HH', obtained from motor perseveration tests	Number written in unit time
	Speed of tracing, obtained from preliminary part of mirror drawing test	Distance covered in unit time on circumference of star
Introversion	Minnesota Personal Traits Rating Scale II	Self-rating
'*W*'	Webb's questionnaire	Self-rating
Surgency	Cattell's questionnaire from Webb	Self-rating
Intelligence	Terman-Merrill revision of the Stanford Binet (1937)	M.A. and I.Q.

II. ADMINISTRATION AND SCORING OF TESTS

(1) *Dark adaptation.* A special apparatus was constructed; a light-proof hood was tied over the subject's head, from which he looked down a tube to a screen on which four letters were faintly illuminated. The illumination was provided by four electric lamps within the box, whose brightness was reduced by a series of screens, and diffused by ground-glass slides. The time required to read the four letters was taken as the measure of the rate of dark adaptation. The areas of the four letters were equal. Rate of dark adaptation was used by Wiersma as a test of 'secondary function'.

(2) *After-image.* A green cross on a grey ground was exposed for periods of 15, 30, and 45 sec. After each exposure the subject was required to look at a silvered screen, and to record appearances and disappearances of the image by pressing a button. Responses were recorded on a kymograph. The score was the sum of three response times. The duration of after-images has not previously been used as a measure of sensory perseveration, but it should serve as a valid measure if there is any validity in the hypothesis of stable individual differences in 'secondary function' or sensory inertia.

(3) *Mirror drawing.* Subjects were instructed to trace (by direct vision) the outline of a six-pointed star as quickly as possible for 30 sec. This was repeated five times. The star was then traced five times by mirror drawing; then once more by direct vision.

Responses were scored by a stencil divided into 108 squares, score being the number of squares touched by the subject's pencil; if the line passed wide of a square, this square was not counted. Three measurements were obtained from this experiment: (*a*) Speed: distance covered in fifth trial by direct vision. (*b*) Motor perseveration (creative effort): total number of squares touched in five trials by mirror drawing. Mirror drawing involves the overcoming of established motor habits in the same sort of way as the motor perseveration tests of Stephenson and Cattell. (*c*) Motor perseveration (alternation): ratio of score in fifth direct trial to final direct trial given immediately after mirror drawing. The theory of this measurement is that the practice of mirror drawing might interfere with the normal performance, and that individual differences in such interference would measure differences in perseveration. In preliminary experiments many subjects reported the experience of such interference.

(4) *Motor perseveration tests.* The four tests used were taken from Cattell. In each test the eight trials were given as follows:

(i) and (iii) First or normal activity.

(ii) and (iv) Second or conflicting activity.

(v), (vi), (vii) and (viii) First and second activities in alternation.

Each trial was 30 sec.

Only two of these, viz. '234' written normally and with reverse stroke, and 'H' written normally and horizontally, were suitable for measuring perseveration of the 'creative effort' type. The other two did not involve abnormal activities, and could only be used for the 'alternation' measurement. The 'creative effort' type was scored in the normal way by the ratio of normal to abnormal activity, measured by number achieved in unit time. For the 'alternating' type, Cattell used the formula

$$p = \frac{(\mathrm{i})+(\mathrm{ii})+(\mathrm{iii})+(\mathrm{iv})}{(\mathrm{v})+(\mathrm{vi})+(\mathrm{vii})+(\mathrm{viii})}.$$

As pointed out by Darroch (1938) this score is influenced by individual differences in the ratio of forward to backward activities, apart altogether from differences in the difficulty of alternation. It thus measures an indeterminate mixture of the 'creative effort' and 'alternation' types of perseveration. Our formula was designed to meet this objection, and is derived as follows: Let us designate the number of responses achieved in the first activity by N_x ($N_x = (\mathrm{i})+(\mathrm{iii})$). N_y is the number of responses achieved in the contrasting activity ($N_y = (\mathrm{ii})+(\mathrm{iv})$; and N_{xy} is the number of responses achieved in the alternating activity ($=(\mathrm{v})+(\mathrm{vi})+(\mathrm{vii})+(\mathrm{viii})$). If no interference occurred, and the rate of each activity was constant, then N_{xy} would be equal to $\frac{4N_xN_y}{N_x+N_y}$. Perseveration (or interference) may then be measured by the difference between the predicted and the actual value of N_{xy}, i.e. by $p=\frac{4N_xN_y}{N_x+N_y}-N_{xy}$. It would perhaps be better to use the ratio rather than the difference, but since our subjects were all normal adults this was of little importance. This score is sometimes in fact negative, because the practice effect is sometimes greater than the interference. In future experiments on this type of perseveration it would be advisable to give extensive practice in both types of independent activity before giving the crucial trials of alternation.

We also used parts of these tests to obtain measurements of speed.

(5) *Tongue twisters.* Three sentences were presented on cards, each to be repeated three times aloud as fast as possible; if the subject failed more than twelve times on one sentence, he passed on to the next. The sentences were:

"She sells sea-shells by the sea shore."

"The Leith police dismisseth us."

"You knew unique New York."

This is a motor interference test, involving a conflict between what one has just said, and what one is trying to say. If then there are stable individual differences in the 'alternation' type of motor perseveration, they should show themselves also in these responses. We recorded time, errors, and number of repetitions required for success. Of these, the third was selected as the most suitable measure.

(6) *Colour naming test.* The words *purple, red, yellow, green, orange, blue,* were printed each four times in varying order on a card 24 × 18 in. Four colours of ink were used—red, blue, green and orange. Subjects were required to name the colours of the inks in which the words were printed. In no case was a word printed in the colour which it named. Time and errors were recorded. Since errors were not numerous, interference was measured by time. The task is analogous to the 'creative effort' type of motor test.

(7) *Free association test.* A free association list of fifty words selected from Cattell (1936) was given twice in immediate succession. It was found necessary to explain that the words would be repeated, and that the subject was to continue to respond with the first word that occurred to him, whether it was the same as, or different from, before. Perseveration was measured by the number of occasions on which the same response was given at the first and second trials. According to the notions of G. E. Müller, subjects with high associative perseveration should repeat their first responses more often.

Median reaction time was used as a measure of fluency. The median rather than the mean was used in order to eliminate the effects of a few long delays due to emotional disturbance.

(8) *Continuous association.* Subjects were required to associate for 3 min. to an initial stimulus word. Two measurements of associative perseveration were obtained from this test. One was the number of occasions on which a word was repeated in the course of the associations. This is another example of G. E. Müller's type of perseveration. The other measurement was based on the length of the trends discoverable within the associative series. For example, if the stimulus word were '*winter*', and the subsequent responses were '*spring, autumn, season, seasoning, mint, parsley*', we should regard the series '*spring, autumn, season*' as a single trend, because each of these words has a natural link with the initial word '*winter*'. But the word '*seasoning*' begins a new trend, no longer directly linked with the original stimulus. The analysis of these trends was made with the assistance of the subjects, but even so was somewhat subjective. Mean length of trends was calculated for each subject. The number of words written down in 3 min. was used as a measure of fluency.

(9) *Fluency tests.* These tests were based on the work of Studman (1935). In preliminary trials it was found that Studman's verbal tests were too easy for our rather intelligent group of adults (mean Stanford-Binet I.Q. 130), so that for our subjects her tests measured little besides speed of writing. In our tests four concepts were employed: 'Words beginning with *Y*', 'things run by electricity', 'games', 'wild animals'. Subjects were required to write down as many of each class as they could think of in 2 min. We also obtained the

number of responses to three of Rorschach's ink-blots (2 min. each); and the number of responses to Studman's three pictures (2 min.). These last are simple drawings of a library, a street, and a tree. There is a cross marked on each picture, and subjects were required to write down as many things as they could think of, which might lie where the cross was.

In adapting these tests we acted on the assumption that fluency is to be measured by the frequency and length of blockages in association, and that the time given to each test must be long enough for blockages of some length to occur. Controlled association scores must not be affected by knowledge; every subject must be familiar with far more examples of the class of objects required than he is capable of enumerating in 2 min.

(10) *The 'W' factor.* A self-rating for 'W' traits on a five point scale was obtained from each subject. The definitions of each trait as in Cattell (1933) were given to the subjects; and difficulties of interpretation were discussed with them individually. The traits rated were:

Persistent—Changeable
Energetic—Inert
Tactful—Tactless
Mature—Wilful
Pedantic—Casual
Not Hedonistic—Sensuous
Confident—Anxious
Kind on Principle—Not So

(11) *Surgency.* Self-ratings were obtained in the same way, also using Cattell's definitions. Traits rated were:

Cheerful—Gloomy
Natural—Formal
Sociable—Unsociable
Adaptable—Conservative
Quick of Apprehension—Slow of Apprehension
Forward—Shy
Co-operative—Distrustful
Original—Banal
Expressive—Inhibited
Placid—Given to Mental Conflict
Sophisticated—Naïve

In view of the well-known unreliability of self-ratings, no great reliance can be placed on these tests, though we hoped that the high intelligence of our group would give some value to their judgements. The method used by Webb (1915) and Cattell (1933) by which all subjects were rated by the same small group of judges, who knew all of them intimately, was inapplicable to our subjects, who did not form a highly integrated group.

III. Results

(1) *Perseveration tests*

Table 1 shows the intercorrelations of fifteen supposed measures of perseveration. (1) and (2) are measures of sensory perseveration; (3)–(5) of motor perseveration ('creative effort' type), here called 'motor perseveration *A*'; (6)–(11) are tests of motor perseveration ('alternation' or 'interference' type), here called 'motor perseveration *B*'; (12)–(15) are tests of associative perseveration.

In all tests except (3) (the mirror-drawing test of creative effort), a high score indicates a high degree of perseveration. In order to bring this test into line with the others, we have used the reciprocals of the raw scores in (3) for correlating.

The probable errors of the coefficients shown in Table 1 are as follows:

r	P.E. r	r	P.E. r
From 0 to ±0·07	±0·10	From ±0·47 to ±0·57	±0·07
From ±0·08 to ±0·33	±0·09	From ±0·58 to ±0·65	±0·06
From ±0·34 to ±0·46	±0·08		

Except where otherwise stated, these are the probable errors of all coefficients of correlation given in this paper.

Table 1. *Correlations between fifteen measures of perseveration*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	—	−0·01	0·01	0·01	0·06	0·01	0·05	0·17	0·01	−0·28	0·16	−0·12	0·07	−0·09	−0·12
2		—	0·01	0·26	0·13	0·00	0·17	0·10	−0·03	0·00	0·01	0·31	0·01	−0·18	0·01
3			—	−0·04	−0·22	−0·06	−0·15	0·03	0·00	0·12	−0·26	−0·16	0·22	0·00	0·01
4				—	0·32	−0·27	−0·05	−0·02	−0·03	−0·06	−0·01	0·35	−0·20	−0·19	0·13
5					—	0·03	−0·07	−0·14	−0·09	−0·08	0·06	−0·12	−0·04	0·09	−0·12
6						—	0·35	0·04	0·19	−0·02	0·18	−0·03	−0·03	0·00	−0·04
7							—	−0·11	0·21	−0·08	0·10	−0·19	−0·18	−0·02	0·01
8								—	0·40	0·11	0·26	0·12	0·09	0·05	0·11
9									—	−0·16	0·19	−0·14	0·04	−0·15	−0·01
10										—	−0·19	0·12	0·24	0·12	−0·10
11											—	−0·06	0·29	0·21	−0·06
12												—	0·06	−0·29	0·07
13													—	0·05	−0·01
14														—	−0·05
15															—

1. Dark adaptation
2. After-image
3. Mirror drawing (distance)
4. Motor perseveration '*A*' (234)
5. Motor perseveration '*A*' (HH...H H...)
6. Motor perseveration '*B*' (234)
7. Motor perseveration '*B*' (ABCDabcd)
8. Motor perseveration '*B*' (H H)
9. Motor perseveration '*B*' ($\pm\infty$)
10. Mirror drawing (interference)
11. Tongue twisters
12. Colour naming
13. Word association (repetitions)
14. Continuous association (repetitions)
15. Continuous association (length of trend)

The correlation between the two measures of sensory perseveration is $-0{\cdot}01 \pm 0{\cdot}10$. The absence of any correlation between these two tests suggests that the hypothesis of Gross, Heymans and Wiersma concerning 'secondary function' is false. Wiersma's tests were applied mainly to the insane, and it is obvious that introspections by the insane, unsupported by any objective check, are not of great value. Our method of measuring the rate of dark adaptation by means of reading letters has the merit of being objective. Wiersma also claimed to measure perseveration by the rate of flicker-fusion. Wynn-Jones (1928), who repeated his work, did not confirm his results. On the whole, the evidence for a general factor of sensory perseveration is not good. The evidence of Biesheuvel's (1938) experiments on the rate of flicker-fusion, and their relation to temperament, is indecisive.

Our evidence for the 'creative effort' type of motor perseveration is inconclusive; while the mirror drawing test is clearly unrelated to the others, it is possible that there is a significant relation between (4) and (5).

In tests (6)–(11) (motor perseveration of the 'alternation' or 'interference' type), an examination of the table reveals that (10) (mirror drawing interference) is unrelated to the others. If, however, this test is omitted, the others have a mean intercorrelation of $0{\cdot}181 \pm 0{\cdot}030$, which reveals a genuine though very small common factor.

Tests (12)–(15) measure associative perseveration. These tests have a mean intercorrelation of $-0{\cdot}048 \pm 0{\cdot}039$. We may therefore conclude that so far as our evidence goes this trait does not exist. Lankes (1914), Pinard (1932), and Cattell (1933) all failed to obtain a reliable measure of this tendency.

We consider next the interrelations of these supposed forms of perseveration. Table 2, derived from Table 1, shows the relations between the four types of perseveration, obtained by calculating the means of intercorrelations of subtests in each group.

Table 2. *Mean intercorrelations of four groups of subtests*

	Sensory perseveration	Motor perseveration '*A*'	Motor perseveration '*B*'	Associative perseveration
Sensory perseveration	—	0·080	0·019	−0·014
Motor perseveration '*A*'	±0·039	—	−0·058	−0·003
Motor perseveration '*B*'	±0·030	±0·022	—	−0·002
Associative perseveration	±0·034	±0·028	±0·019	—

It is plain that there is no significant correlation between any of these groups. The absence of correlation between the two types of motor perseveration is particularly interesting, because previous investigators have used methods of scoring which combine these two in unknown proportions (e.g. Stephenson (1934), Rangachar (1932), Cattell (1935)). This is probably one of the reasons for the low reliability of perseveration tests. Of the eighty correlations between tests belonging to different groups, thirty-nine are positive, forty-one negative. Their mean value is $0{\cdot}005 \pm 0{\cdot}011$. Their standard deviation is 0·131. If the true value of the mean were 0, the theoretical standard deviation would be $0{\cdot}141 \pm 0{\cdot}008$. It therefore appears probable that the true value of these correlations is precisely 0. Our results thus give no support to Spearman's General Law of Inertia. Of the four types of perseveration discussed in the literature, two appear to be non-existent, and the remaining two appear to be unrelated to one another.

(2) *Fluency tests*

Correlations between measures of fluency contrast sharply with those between measures of perseveration. Table 3 presents correlations between five supposed tests of fluency. All the correlations have the expected signs, and only one is less than three times its probable error. Correlations with free association (reaction time) are naturally negative, because shortness of time (hence low scores) indicates high fluency, while in the other tests high score indicates high fluency.

Table 3. *Intercorrelations of five tests of fluency*

	1	2	3	4	5
1. Controlled association	—	0·46	0·50	0·34	−0·30
2. Ink-blot interpretation		—	0·58	0·43	−0·29
3. Picture interpretation			—	0·40	−0·10
4. Continuous association				—	−0·51
5. Word association (median time)					—

(3) *Relations to supposed factors of temperament and personality*

In order to consider the relations of the qualities measured to the various measures of temperament employed, we reduced the scores in motor perseveration ('alternation' type) to standard scores, and summed. Scores in speed and fluency were also reduced to single figures, by weighting the scores to give approximately equal standard deviations, and summing. No combined scores were obtained for the other types of perseveration, because, unlike Cattell (1933), we can see no value in composite scores which are obtained from tests virtually uncorrelated with one another. Table 4 gives the correlations of motor perseveration with other qualities measured.

Table 4

	Fluency	'*W*'	Introversion	Speed	Intelligence	Surgency
Motor perseveration '*B*' (alternation)	0·27	−0·08	−0·03	0·02	0·41	0·04

Most previous investigators (e.g. Pinard, 1932) have found no relation between introversion and perseveration, so our result is not surprising. The significant correlation with intelligence agrees with Pinard (1932), but is contrary to the results of most other investigators. It must be remembered, however, that previous investigators have not obtained a pure measure of this type of perseveration.

Cattell (1933) asserts that high and low perseverators have low '*W*', and moderate perseverators have high '*W*'. In order to test this, we correlated scores in '*W*' with the absolute values of standard scores in motor perseveration. If Cattell's view is confirmed, the correlation should be negative. We find a correlation of $0{\cdot}15 \pm 0{\cdot}09$.

Correlations of fluency with other qualities are given in Table 5.

Table 5

	'*W*'	Introversion	Speed	Intelligence	Surgency	Motor-perseveration '*B*'
Fluency	0·12	−0·24	0·28	0·28	0·09	0·27

The correlations of fluency with speed and intelligence are similar to those found by Cattell and Studman. The absence of significant correlation with surgency fails to confirm

their results, though, as we have pointed out, we can place no great reliance on our self-ratings. The negative correlation with introversion fits in with Cattell's view of fluency.

IV. Summary

Fifteen measures of various aspects of perseveration were obtained from a group of fifty educated adults. The results fail to confirm the hypothesis of a law of inertia, or diffused tendency to perseveration, but give some evidence of a general factor of interference in alternating motor tasks. The existence of a general factor of speed of association, or fluency, is strongly confirmed. Relations with intelligence and temperamental qualities are also considered.

REFERENCES

Biesheuvel, S. (1938). "The measurement of the threshold for flicker and its value as a perseveration test." *Brit. J. Psychol.* xxix, 27.

Cattell, R. B. (1933). "Temperament tests." *Brit. J. Psychol.* xxiii, 308; xxiv, 20.

Cattell, R. B. (1935). "On the measurement of perseveration." *Brit. J. Educ. Psychol.* v, 76.

Cattell, R. B. (1936). *A Guide to Mental Testing.* University of London Press.

Darroch, J. (1938). "Variations in the score of a motor perseveration test." *Brit. J. Psychol.* xxviii, 248.

Gross, O. (1902). *Die Cerebrale Sekundarfunktion.* Leipzig.

Lankes, W. (1914). "Perseveration." *Brit. J. Psychol.* vii, 387.

Müller & Pilzecker (1900). *Z. Psychol.* Erg.-Bd. i, Chap. 3.

Pinard, J. (1932). "Tests of perseveration." *Brit. J. Psychol.* xxiii, 5, 114.

Rangachar, C. (1932). "Differences in perseveration among English and Jewish Boys." *Brit. J. Educ. Psychol.* ii, 199.

Spearman, C. (1926). *The Abilities of Man.* London.

Stephenson, W. (1934). "An introduction to so-called motor perseveration tests." *Brit. J. Educ. Psychol.* iv, 186.

Stroop, J. R. (1935). "Studies of interference in serial verbal reactions." *J. Exp. Psychol.* xviii, 643.

Studman, L. G. (1935). "Studies in experimental psychiatry. V." *J. Ment. Sci.* lxxxi, 107.

Webb, E. (1915). "Character and intelligence." *Brit. J. Psychol., Monogr. Suppl.* i, No. 3.

Wiersma, E. (1906). "Die Sekundarfunktion bei Psychosen." *J. Psychol. Neurol.* viii, 1.

Wynn-Jones, Ll. (1928). "An investigation into the significance of perseveration." *J. Ment. Sci.* lxxiv, 653.

(*Manuscript received* 23 *December* 1942)

NOTE ON THE NUMBER CONCEPTION IN ANIMAL PSYCHOLOGY

By D. H. SALMAN

I. *Introduction* (p. 209). II. *Unsatisfactory methods of study* (p. 209). III. *Description and classification of 'counting' behaviour* (p. 210). IV. *The abstract concept of number* (p. 211). V. *Training in abstraction of number* (p. 212). VI. *Suggestions for future procedure* (p. 216). *References* (p. 218).

I. Introduction

H. Honigmann's (1942*a*) recent review and able discussion of the studies so far published on "The number conception in animal psychology" suggest several theoretical remarks that may be of some interest in view of future research both on this problem, and in the wider field of general animal psychology. A preliminary paragraph will first discuss one particular method which seems inadequate. The central problem will then be considered in detail, with its essential theoretical implications.

II. Unsatisfactory methods of study

Several authors describe the simultaneous training of their animals to carry out several different tasks (Sadovinkova, 1923; Arndt, 1939; Marold, 1939; Schiemann, 1940). In the most favourable case, Schiemann taught his jackdaws to take 2, 3, 4 or 5 pieces of bait, a different colour being shown in each case to signal which particular response was expected. Many authors seem to consider that such a multiple-task training has some relation with the number problem. But this is surely a fallacy. The successful accomplishment of such multiple tasks illustrates the animal's capacities for complex differential and simultaneous training, but not in any way its ability to count.

This can best be shown by the following analysis. All animals possess the ability to respond in a great many different ways to different situations. Schiemann's jackdaw, for instance, would no doubt have reacted in a different way to, say, a mealworm, a grape, a nut, a cherry and a robin's egg; and the list could be lengthened indefinitely. But this does not in any way demonstrate its ability to count to five. Or else an earthworm could be said to count to 150, since the sharp pricking of any one of its segments arouses a differential response. The behaviour shown in Schiemann's experiment differs from that just mentioned in two respects only. First, the primary and spontaneous valences of natural objects (nuts, cherries) were replaced by the secondary and artificial valences acquired by coloured papers as a result of training. This difference makes the whole experiment a measure of the jackdaw's ability to learn. The second difference is that the expected response (viz. the eating of 2, 3, 4 or 5 pieces of bait) is in itself an achievement which connotes some ability to discriminate between 2 and 3 pieces. In this way, one of the constituents of the observed behaviour (i.e. the response) proves by its very existence some capacity for number discrimination. The individual tasks here used have, as such, some relation with the number problem; but the multiple tasks, in so far as *multiple*, have no relation whatever to it.

III. Description and classification of 'counting' behaviour

But what exactly is achieved by the best-trained animals? To appreciate this some definition would be required as to what 'counting' is, or what can be defined as "a preliminary stage to counting" (Honigmann, 1942*a*, p. 243). More important still, a specific behaviour would have to be described, which would be recognized as proof that the corresponding 'preliminary stage to counting' has actually been achieved. Such a classification of counting behaviours, i.e. functional definitions of counting and its eventual component parts, has remained largely implicit in the relevant literature. Most authors, if they do mention such theoretical principles, only refer to them hurriedly to justify the (alleged) significance of their own results.

The actual research procedure of the various authors, however, gives valuable clues concerning their unuttered principles; and the important work already accomplished gives a fairly clear indication as to what is generally intended by 'counting behaviour'. This we shall now attempt to formulate explicitly; the validity of these implicit notions will then be critically examined.

The trained animal reacts in a specific manner to a specific situation: and somewhere, in the stimulus or in the response, a number occurs that contributes to the specific character of the behaviour. Now it is clear that the animal's behaviour is a sensori-motor complex, integrated in space and time. Any change in the conditions may (and often will) bring about a change in the response: and this will require added training, in order to eliminate the troubling feature while retaining the specific 'number-character' of the intended behaviour. In this way, the 'number-character', at first hidden in a mass of irrelevant and indifferent detail, will become by progressive eliminations an increasingly important determining factor of the animal's behaviour. When the process is complete, the 'number-character' will be dominant. And the exact nature of the non-eliminated remainder will then determine the significance of the response, and allow an estimate of the 'number-character' of the observed behaviour.

Now in actual research, some variations are thought irrelevant to the number problem, and are not considered worth while eliminating. Hens trained to eat the odd grains of an equally spaced row, leaving the even ones untouched, were quite unable to respond adequately when the normal row was presented far from the usual place, or on a background of a different colour (Honigmann, 1942*b*). Monkeys trained to distinguish series of sounds lost all their abilities when the quality of the sound was changed: an electromagnetic sound-hammer instead of blows on a metal container (Woodrow, 1929). These and similar observations (Köhler, 1918; Fischel, 1926) are significant as revealing the importance of apparently unessential components in the specification of an animal's behaviour. But they can, with some apparent justification, be provisionally disregarded as extraneous to the 'number concept'.

Other aspects of the animal's integrated behaviour have been considered more important, and great care and ingenuity has been spent in the elimination of these factors, in order (apparently) to isolate the 'number-character' more completely. Thus the remarkable training of monkeys by Gallis (1932) and hens by Giltay (1933) was shown by Bierens De Haan (1935, 1936*a*, *b*) to rest on the learning of certain temporal rhythms of action. This was considered as alien to the 'number-concept', thus disproving Giltay's and Gallis's claims to have observed 'counting'. We therefore find Koehler and Wachholz

(1936), Arndt (1939) and Schiemann (1940) devising clever experiments to eliminate every possible time-pattern, be it in the presentation of the stimulus, in pecking rhythm or in swallowing rhythm.

The same authors also took great care to eliminate figural help. For it was found that the discrimination of 3 and 4 peas was much facilitated when they were spatially grouped as .·. and ...., and similarly for other numbers and groups. Consequently, extensive variations were undertaken in the training conditions to exclude all figural element, thus again (apparently) giving a more definite 'number-character' to the animal's behaviour.

But what exactly is accomplished by these careful and thoughtfully designed experiments? The whole process of differential training, in which many conditions are constantly varied to ensure the elimination of the unwanted factors, might be called one of growing abstraction; or, far better, a process of simplification and progressive schematization. From the first, the animal took 4 peas. But initially it relied on the figural configuration and the motor pattern of its rhythmic pecking and swallowing. When these cues were eliminated by an adequate variation of the experiment, a simplified and more schematic character of the total situation became specific. For time-patterns and spatial configuration had become irrelevant, meaningless; better still, they were now deprived of valence. The positive valence which, when the training began, was associated with the complex situation as a whole, finally remained attached to whatever subsists as non-eliminated residue of the varying conditions of the training.

A question now arises: how far can one go in this process of simplification? And if one must stop somewhere, what is the structure of the 'non-eliminated residue', and the resulting 'number-character' of the corresponding behaviour? The answer to these questions will appear with greater clarity if the extreme case of total abstraction is first considered.

IV. The abstract concept of number

The completely abstract concept of number has been observed, so far, only in human beings. Functionally defined, it results from a training, which young children actually undergo, and in which stimulus and response in some way include the required number. Variations of the experiment (unsystematic variations in general) finally eliminate all extraneous factors, just as in the case of animal training. What then remains? A completely abstract pattern; or, in the equivalent functional definition, a pattern which is exemplified in many situations having no concrete characters whatever in common. The number 4 may mean four dots on a background in any configuration: and so figural elements have been eliminated, as in the case of pigeons. It may mean four similar actions, in any space-time interval: thus eliminating rhythmic patterns, as in the case of monkeys, hens, jackdaws and parrakeets. But the child can also be taught to eliminate all particular colours, sizes, locations: which animals have found well-nigh impossible. It can eliminate all particular qualities, and recognize four peals of a bell, four blasts of a trumpet, four shots of a gun, four blows, four kisses. It will respond to four flashes of a signalling lamp (extended in time), four fingers touching its hand (extended in space), four voices singing together (plural in quality). The possible variations are indefinite. But the essential point is that two examples may have nothing whatever in common. Or, taken from another point of view, the 'non-eliminated residue' that remains when all possible conditions of the experiment have been varied is non-representable, non-

imaginable: an abstract pattern that can perhaps be conceived, but that can surely not be perceived or sensed as such.

The 'human' or abstract concept of number will therefore be achieved if, and when, the trained animal responds to 'four' or 'five' in any of the utterly diverse concrete circumstances above suggested; and if, when confronted with a new field of sense-patterns, it can transfer to it the 'four-pattern' acquired elsewhere. Such would be, for instance, the recognition of four blasts of a whistle, after training to four dots on a paper or four grains to be eaten. If the abstract concept of number is ever to be achieved, such transfers must occur. For it is clearly impossible to eliminate by training the qualitative properties of all possible concrete situations.

As long as this abstract 'number-concept' remains the ideal, the successive eliminations of figural elements, time-patterns, pecking rhythms, etc., are thoroughly justified. For all these factors are concrete determinations, that cannot consistently be retained where a humanly abstract number-pattern is required. And such seems to have been the (implicit) attitude of all the above-named authors. It has seemed essential to formulate this fundamental assumption explicitly, since it explains the experimental method hitherto used.

If, however, this ideal were thought to be beyond the possibilities of any animal, a far different procedure would have to be adopted. But even then the abstract 'number-pattern', and the corresponding behaviour, would deserve to be studied in some detail, as suggesting a convenient model, to which lesser accomplishments may usefully be compared. And it will be used in this way, in the study of the actually observed achievements of animals, to which we may now return.

V. Training in abstraction of number

It is easy to define the training which corresponds to the abstract 'human' number-concept, for in that case all particular qualities must be eliminated by a suitable variation of experimental conditions. If anything less ambitious is attempted, many concrete specifications will remain. How far, then, can one go? How far should one go? And what will be the nature of the 'non-eliminated residue' that decides the meaning of the experiment?

How far can one go? That is a problem to be settled by experimental research. Granted sufficient training, and a well-adapted technique, it has been found possible to eliminate relative location, spatial configuration, and all temporal rhythm-patterns. What then remains?

First, a great number of 'irrelevant' specifications, such as the colour of background and apparatus, their absolute size, the location of the experiment, etc. These factors are considered irrelevant to the number problem. But they are certainly not irrelevant to the experiment as a whole, for a variation in any one of them will usually lead to an inadequate response, proving that for the animal they are a significant part of the stimulus to which it reacts.

The other, and more essential remainder, is a certain sensory pattern. This, it is important to note, is always of a definite kind: visual, or sensori-motor, or auditory, etc. A training is always bound to one of these sensory fields; and an acquired behaviour in one of them gives no help towards a similar pattern in any other. No animal, in other words, has shown any capacity for the transfer of a number-pattern from one qualitative

field to another. Indeed, the reverse has been observed. The most gifted of Schiemann's jackdaws found it hard to associate 2 seen dots with the eating of 2 pieces of bait, and 4 dots with 4 pieces of bait; while all the birds easily learned to associate the eating of 2, 3, 4 or 5 pieces of bait with different colours.

Within a special field of sense qualities, the number-pattern seems fairly indefinite, since figural factors could be eliminated from all visual number-perceptions, and temporal rhythm from sensori-motor number-behaviours. Some pattern or organization of the sense qualities must, however, remain, since the animal in some way discriminates between 4 and 5.

Patterns of this type have long been known and studied in human psychology, by Brunswik, Holaday and their collaborators (cf. Brunswik, 1934). The subjects were asked, for instance, to appreciate the number of coins on a plate, or stamps on a blotter. They were not of course allowed to count, and the numbers which had been chosen were large enough to exclude all spontaneous recognition of familiar patterns such as 3, 4, or 5. At the same time, the conditions of the experiment, such as colour, size, background and spatial configuration, were extensively varied, not to eliminate these unescapable characters, but in order to appreciate their relative importance. Now this is exactly the situation confronting Arndt's pigeons or Schiemann's jackdaws, when trying to recognize the number of dots on a background; although the training could be dispensed with in view of the human subject's preliminary (and transferable) conditioning in the use of number-patterns.

It was found by Holaday that such perceptions are of a complex nature; they represent no simple function, but are the result of an integrated whole in which most of the 'irrelevant' factors, such as brightness, colour, shape and size, play as essential a part as the real (abstract) number of objects. Moreover, such a function is characteristic of the sensory field in which it occurs, and cannot be transferred from one such field into another. The same complex type of function would also be found in the problem of size-constancy, which has been extensively studied both in animals (Köhler, 1915, 1917; Götz, 1926) and man (Brunswik, 1933, 1934; Holaday, 1933; cf. Frank, 1928; Thouless, 1931, 1932; Koffka, 1935). A comparative study of this material should prove illuminating, but lies beyond the intention of the present paper.

What, then, is the 'number-character' of the behaviour so far achieved by dint of patient training? Not, of course, an abstract number-pattern of the human type defined above. Nor, and this is essential, the same approximation to an abstract number-pattern, no one and identical 'preliminary stage to counting' that would be common to all behaviours. The monkeys trained by Kuroda (1931) and Woodrow (1929) responded to sound-patterns, which were combined with a large number of specifying qualities, many of which were significant. This is one possible type of non-abstract number-pattern. Koehler and his collaborators (1935, 1936), Arndt, Marold and Schiemann trained various birds to repeat an action x times: e.g. to eat 5 grains from a heap containing many more. This is another, and quite different type of non-abstract number-pattern, which is of a fundamentally sensori-motor kind, although it of course includes many local qualitative specifications. Finches and pigeons were trained by Fischel, and pigeons again by Koehler *et al.* to discriminate between 3 and 1 grains, or 4 and 2, 5 and 3, etc. Here we have, once again, a different, and this time visual, type of number-pattern, from which figural elements may have been eliminated, but which still includes many significant

size, colour and location factors. Other experiments include more complex specifications. Marold's budgerigars, for instance, learned to eat 2, 3 or 4 grains when shown a black cardboard with a yellow border, a plain black cardboard, or a yellow one with a black border, respectively. Arndt's pigeons had earlier performed tasks of the same kind. This is no doubt a non-abstract sensori-motor number pattern, but of a type quite distinct from that above mentioned, since it includes significant specifications of a visual nature in which both colour and shape are involved. In some cases, finally, two of the above non-abstract number patterns are combined, together with local specifications. Such was the case when one of Schiemann's jackdaws learnt to eat 2 mealworms when shown 2 dots, and 4 mealworms when shown 4 dots. This is yet another type of non-abstract number-pattern, in which visual and motor patterns are combined in still another way.

These behaviours, I submit, cannot be lumped together as exemplifying a common number-character. The first three should be classified as training in sound patterns, non-rhythmic motor patterns and non-figural visual patterns, respectively. Words fail to express, even approximately, the nature of the others, unless they be called 'non-rhythmic motor *cum* figural-colour' pattern for the one, and 'non-rhythmic motor *cum* non-figural-visual' for the other. For an animal, then, these behaviours have nothing in common. A man might recognize in each a same abstract number-pattern. But that is a human point of view, not one that is adequate or significant in animal psychology.[1]

Even similar trainings of animals of the same species are difficult to compare. For local conditions are bound to vary, and all have been found significant. Nor is it possible to retrain an animal in all possible locations, with all possible backgrounds, with instruments of all possible sizes and colours, in order to eliminate these factors. Finally, one must come to the conclusion that an animal's acquired behaviour can only be defined by the conditions, by *all* the conditions, of its training. To call such a result a 'number-behaviour' is surely unsound. To call it a non-rhythmic motor pattern, or a non-figural visual pattern, as the case may be, is more accurate; but it may still be seriously misleading, if the other non-eliminated qualifying factors of the training are not explicitly stated. The behaviour is adequately defined by the training from which it derives; to designate it in any other way is always difficult; to label it with a notion foreign to animal psychology such as 'number' is an error in method that should be studiously avoided.

It is interesting to note that a similar fallacy has been found to occur in the development of human psychology. Preliminary training is here replaced by the acquisition of language, so that behaviour is directly observable in the performance of so-called intelligence tests. When these were first used, they were often given the names of 'faculties' or mental traits to which they were supposed to correspond. The exact statistical analysis of the Spearman-Thorndike-Thurstone school has conclusively shown how ill-founded were these early speculations. Tests which seemed very similar, and were thought to

[1] O. Koehler partly realized this in the particular field he was studying (Koehler & Wachholz (1936)). For he explicitly discriminates between a 'basic ability no. 1' which is a choice between different numbers, and a 'basic ability no. 2', which consists in the eating of a definite number of grains from a pile. He manifestly saw the fundamental difference between what we have called a non-figural visual pattern and a non-rhythmic motor pattern. Actually, basic ability no. 1 also includes a choice, while basic ability no. 2 consists in the repetition of a single action; and that is another essential difference. But the significance of this discrimination is not further stressed, nor does Koehler consider the other possible types of non-abstract number-patterns, even those that actually occur in his experiments.

derive from the same psychological trait, were shown to have very low correlations, while other tests which had seemed very dissimilar proved to be highly correlated. The uncritical use of introspective notions, in other words, was illegitimate, and the classifications derived from them were totally inadequate. The immediate meaning of a test is given by its very nature. What is tested is the ability to respond to that particular test, that and no more; just as, in animal psychology, an acquired behaviour proves a capacity to respond to the training conditions, that and no more.

Statistical analysis eventually led to the isolation of the factorial 'abilities' which operate, in various combinations and proportions, in the performance of tests. But this was achieved by the analysis of the experimental data, and not by the introduction of extraneous categories. Similarly, a careful analysis of animal behaviour can lead to a more systematic knowledge of the factors involved. But this will never be accomplished as long as notions are used that are foreign to the experimental data, such as the 'number-concept'.

The pursuit of an abstract number-pattern in animal behaviour has led to many artificial situations. Animals have been trained to extraordinary sensory configurations, the like of which they would certainly never meet in wild life. It is significant that an enormous amount of training was required even in the simplest of cases. For instance, to train a budgerigar to eat 6 grains from a heap (of 10–15 grains usually) more than 600 experiments were needed with the same animal, while severe punishment was inflicted in case of a wrong choice (Marold). Honigmann fittingly remarks (1942*a*, p. 332): "Thus the first success was visible during the seventh hundred (Experiments 601–700), and here still 43 % mistakes were made." Such conditioning could not conceivably occur in wild life, and certainly plays no part in normal animal behaviour.

In the attempt to isolate an abstract 'number-concept', perceptual patterns have been forcibly pried apart that in normal life are always combined. Thus, all the authors insist on the mutual interaction of 'number', and figural elements (Fischel; Koehler, Müller and Wachholz; Koehler and Wachholz; Arndt). Schiemann's most intelligent jackdaw, in what was for it a difficult task (2 dots or 4 as indication that 2 or 4 pieces of bait were to be eaten) could only achieve 'a certain independence' of figural help. Non-rhythmic time-patterns also are surely very artificial for a bird.

Such conditions are not only artificial, they are ultimately ineffective. For if swallowing 4 pieces of bait is to be discriminated from eating 5, some (albeit non-rhythmic) type of motor pattern must remain. Indeed, its existence has actually been observed by Schiemann. I quote Honigmann's excellent summary (1942*a*, p. 334): "One jackdaw, during a series of trials where the task was to take 5 mealworms out of 4 dishes, had been punished for opening too many dishes. In the following two tests the bird took only 3 and 4 worms respectively, instead of the 5 allowed. During the next trial 6 dishes with 5 mealworms were used in the following way:

(I)	(II)	(I)	( )	(I)	( )
1	2	3	4	5	6

i.e. dish 2 contained 2 mealworms, dishes 4 and 6 were empty and the rest contained 1 mealworm each. The jackdaw, still under the influence of the former punishment, opened only the first 4 dishes and thus obtained only 4 mealworms instead of the 5 permitted. Then the bird, after looking at the punishing device, went back to the door and started again: it nodded once at dish 1 (open and empty), nodded twice at dish 2,

again once at dish 3. Then, after a short pause, and without taking any notice of dish 4, the bird opened dish 5, and took the missing mealworm. The author says that the bird, as it were, repeated its action with 'intentional movements', and only in the case of the still closed dish 5 actually performed the pecking." Such behaviour clearly shows the existence of a motor pattern, which may include no temporal rhythm, but surely does retain a structure by which the bird discriminates between the finished action and the incompleted one. There still remains, therefore, a mixture of sensori-motor patterns and 'number-character'. Similarly, a group of dots must retain some spatial configuration by which they function as *one* group of (say) 5. And this is a minor but very essential kind of figural element, which is undoubtedly combined with the 'number-character' and contributes, together with absolute size, colour, brightness, etc., to the discrimination (cf. Holaday *apud* Brunswik, 1934).

It is clear, therefore, that all these animals have been the victims of very artificial forms of training. This has led to abnormal behaviours, resulting in the isolation of teratologous perceptual and motor patterns. The artefacts thus evolved are peculiar in each case to the specific and highly complex set of conditions from which they derive. The different observed behaviours, and the corresponding perceptual and motor patterns, cannot be ascribed to an identical abstract number-concept; nor can they be reduced to a common non-abstract form, nor even to a small number of comparable types. For the classification thus attempted is foreign to the observed behaviour, and its various elements are no true factors of normal animal psychology.

Such mistaken and ineffective attempts artificially to isolate factors which have no normal independent existence is a fallacy well illustrated in the development of human psychology. In this case, introspective methods were used, and the isolation of so-called 'pure sensation' was the goal to be attained. Or again, in a far different school of thought, a simple two-term reflex arc was the intended atomic factor, which was to be isolated by suitable conditioning. Neither proved an effective means of research, and such analytical procedures were superseded by methods which put greater emphasis on integrated 'wholes' (Köhler, 1929; Lashley, 1929). It will be our task here to discover the 'wholes' that are peculiar to each type of animal, and are therefore significant in its psychological life.

VI. Suggestions for future procedure

The methods hitherto used had some justification as long as it could be hoped to achieve a really abstract number-pattern of the type defined above. If, however, this typically human ideal is abandoned, the violent production of meaningless artefacts will have to be rejected, and the whole of our research on the higher functions of animals reorientated in a different way. This does not mean that the training of animals to carry out discrimination tests should be discontinued. But it does imply that these tests should be used in a manner tending to reveal the animal's peculiar psychology rather than its sufferings when mangled on the procrustean bed of abstract human categories.

The formulation of a general programme of research based on these principles lies beyond the scope of the present note. But an approximate idea of the required attitude may be conveyed by the analysis of some interesting data concerning the number of eggs a bird will lay, and its experimental variation, that has long been classical in ornithology. It is remarkable that none of these observations should even have been mentioned in the extensive literature reviewed by Honigmann, nor indeed in the review itself.

More than 250 years ago, John Ray writes of the lark that "if only one or two eggs are allowed to remain in the nest, the bird will go on to lay for a time indefinite, but if they are three, she will sit. The usual number of eggs in a lark's nest is five." This type of behaviour has been more recently confirmed both in the lark and in other species, such as the blackbird and the long-tailed tit. Kirkman (1911, p. 346) mentions the case of a wry-neck, a species that normally lays seven or eight eggs: by removing the eggs one by one, so that the clutch was never complete, as many as 42 were finally laid, and this was again achieved with the same bird the following year. Witschi (1935) performed the same experiment on swallows, and obtained as many as 50 eggs, with series of 12–19 on consecutive days, instead of the 4 or 5 of the total normal clutch.

This type of behaviour of course only occurs in the species of so-called 'indeterminate layers'; these go on laying until a definite accumulation of eggs releases the incubation behaviour, which in turn leads to a rapid regression of the ovary through degeneration of the larger eggs. All these modifications, including both broodiness and ovarian regression, are controlled by prolactine, a hormone produced by the anterior pituitary (Marshall, 1936; Witschi). These hypophysial gonadotropic functions directly involve the hypothalamus, although the exact nature of the interaction is by no means clear (Brooks, 1940; Uotila, 1940). The whole chain of reactions explains, in outline, how the perception of a definite number of eggs can influence the bird's physiological activity.

The egg-laying behaviour described in the preceding paragraph has a definite number-character, since the bird's response is determined by the number of eggs already in the nest. But so far we have barely more than a field observation, coupled it is true with a fair knowledge of the physiological and biochemical background. If this were now to become the object of judicious experimental research, it might lead to a better comprehension of the number-concept in animal psychology. Such experiments would, however, have several significant characteristics, which it may be useful to formulate explicitly:

(1) The total situation is a natural one, such as normally occurs in wild life. Even the disappearance of an egg through loss or destruction is a common event.

(2) The animal's behaviour is therefore spontaneous. Primary valences exist, in other words, so that all artificial training and conditioning can be eliminated. Nor are punishments required, with their sequel of extraneous complications.

(3) The observer tries to find out how the animal behaves, instead of torturing it into performing what he intends. The elucidation of animal categories is attempted, not their violent reduction to human concepts.

(4) Experimental variations and control tests can and should be used. But their function would not be to eliminate the so-called 'irrelevant' factors. For who is the human experimenter to decide what is or is not relevant to the situation from the animal's point of view? Their function would be to discover the relative importance of factors such as: visual size and shape, colour, brightness, tactile size and shape, temperature, hardness, odour, etc., all of which may play a part in the animal's behaviour.

(5) The contents and the structure of the animal's perception could thus be discovered. Or, if a functional formulation is preferred, the nature and interactions of the various stimuli that together determine the animal's response could be ascertained. And, in this manner, some insight would be gained into the way 'numbers' really occur in a swallow's world.

The total situation of this natural experiment is fully as complex and highly specialized as the most sophisticated laboratory training test. But all the factors are here normal and necessary parts of the experiment: instead of an artificial conditioning, the physiological condition of the laying bird; instead of a meaningless (yet relevant) background and location, the bird's own nest, etc. This is a timely reminder of the very peculiar type of number-pattern that is here being analysed. But *this* pattern can be given an adequate name: the swallow's egg-number-pattern. The same bird surely has very different number-patterns when it responds to several displaying males, to several of her chicks, to several food insects, etc. For, once again, abstract number is not an animal category; the artificial man-produced laboratory number-patterns have no significance in animal behaviour; but real concrete and highly distinctive number-patterns do occur in the worlds of an animal like the swallow. A duck's number-patterns would probably be very different; a lizard's, a fish's, a gorilla's or a spider's even more.

REFERENCES

Arndt, W. (1939). "Abschliessende Versuche zur Frage des Zählvermögens der Haustaube." *Z. Tierpsychol.* III, 88–142.

Bierens de Haan, J. A. (1935). "Zahlbegriff und Handlungsrythmus bei einem Affen." *Zool. Jb.* (Abt. 3), LIV, 267–88.

Bierens de Haan, J. A. (1936*a*). "Notion du nombre et faculté de compter chez les animaux." *J. Psychol. norm. path.* XXXIII, 373–413.

Bierens de Haan, J. A. (1936*b*). "Bildung und Zerstörung von Handlungsrythmen bei einem Schweinsaffen." *Arch. néerl. Zool.* II, 143–59.

Brooks, C. Mc. (1940). "Relation of the hypothalamus to gonadotrophic functions of the hypophysis." *Res. Publ. Ass. Nerv. Ment. Dis.* XX, 525–50.

Brunswik, E. V. (1933). "Die Zugänglichkeit von Gegenständen für die Wahrnehmung." *Arch. ges. Psychol.* LXXXVIII, 377–418.

Brunswik, E. V. (1934). *Wahrnehmung und Gegenstandswelt.*

Fischel, W. (1926). "Haben Vögel ein Zahlengedächniss?" *Z. vergl. Physiol.* IV, 345–69.

Frank, H. (1928). "Die Sehgrössenconstanz bei Kindern." *Psychol. Forsch.* X, 102–6.

Gallis, P. (1932). "Les animaux savent-ils compter?" *Bull. Soc. Sci. Liège*, pp. 82–4.

Giltay, M. (1933). "La notion du nombre chez les oiseaux." *Bull. Soc. Sci. Liège*, pp. 142–7.

Götz, W. (1926). "Experimentelle Untersuchungen zum Problem der Sehgrössenconstanz beim Haushuhn." *Z. Psychol.* XCIX, 247–60.

Holaday, B. E. (1933). "Die Grössenconstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen." *Arch. ges. Psychol.* LXXXVIII, 419–86.

Honigmann, H. (1942*a*). "The number conception in animal psychology." *Biol. Rev.* XVII, 315–37.

Honigmann, H. (1942*b*). "The alternation problem in animal psychology." *J. Exp. Biol.* XIX (summarized in the author's review (1942*a*)).

Kirkman, F. B. (1911). *British Bird Book*, II, 346.

Koehler, O., Müller, O. & Wachholz, R. (1935). "Kann die Taube Anzahlen erfassen?" *Verh. dtsch. Zool. Ges.* pp. 39–54.

Koehler, O. & Wachholz, R. (1936). "Weitere Versuche an die Taube 'Nichtweiss' zur Frage des Zahlvermögens." *Verh. dtsch. Zool. Ges.* pp. 211–36.

Koffka, K. (1935). *Principles of Gestalt Psychology.* London.

Köhler, W. (1915). "Optische Untersuchungen am Schimpansen und am Haushuhn." *Abh. Preuss. Akad. Wiss.* no. 3, pp. 1–139.

Köhler, W. (1917). "Die Farbe der Sehdinge beim Schimpansen und beim Haushuhn." *Z. Psychol.* LXXVII, 248–55.

Köhler, W. (1918). "Nachweis einfacher Strukturfunktionen beim Schimpansen und beim Haushuhn." *Abh. Preuss. Akad. Wiss.* pp. 1–101.

Köhler, W. (1929). *Gestalt Psychology.* London.
Kuroda, R. (1931). "On the counting ability of a monkey (Macacus cynomolgus)." *J. Comp. Psychol.* xii, 171–80.
Lashley, K. S. (1929). *Brain Mechanisms and Intelligence.* Chicago.
Marold, E. (1939). "Versuche an Wellensittichen zur Frage des 'Zähl'-vermögens." *Z. Tierpsychol.* iii, 170–223.
Marshall, F. H. A. (1936). "Sexual periodicity and the causes which determine it." *Philos. Trans.* B, ccxxvi, 432–56.
Sadovinkova, M. (1923). "A study of the behaviour of birds by the multiple-choice method." *J. Comp. Psychol.* iii, 249–82.
Schiemann, K. (1940). "Vom Erlernen unbenannter Anzahlen bei Dohlen." *Z. Tierpsychol.* iii, 292–347.
Thouless, R. H. (1931). "Phenomenal regression to the real object." *Brit. J. Psychol.* xxi, 339–59; xxii, 1–30.
Thouless, R. H. (1932). "Individual differences in phenomenal regression." *Brit. J. Psychol.* xxii, 216–41.
Uotila, U. U. (1940). "Hypothalamic control of anterior pituitary function." *Res. Publ. Ass. Nerv. Ment. Dis.* xx, 580–8.
Witschi, E. (1935). "Seasonal sex-characters in birds and their hormonal control." *Wilson Bull.* xlvii, 177–88.
Woodrow, H. (1929). "Discrimination by the monkey of temporal sequences of varying number of stimuli." *J. Comp. Psychol.* ix, 123–57.

(*Manuscript received* 19 *January* 1943)

TYPES OF LEARNING IN INSECTS AND OTHER ARTHROPODS

By W. H. THORPE, *From the Department of Entomology, Zoological Laboratory, Cambridge*

PART I

I. *Introduction: Definition of terms* (p. 220). II. *Habituation* (p. 225). III. *Habituation in ants and other social Hymenoptera* (p. 226). IV. *Associative learning* (p. 227): (1) *Conditioning* (p. 228); (2) *Trial and error learning*: (*a*) *Simpler types* (p. 228); (*b*) *More complex types* (p. 229).

I. Introduction: definition of terms

There is much in arthropod behaviour, particularly that of insects, which suggests the existence of learning ability of a relatively high order. But in very few instances have the recorded observations been critically considered in relation to the different types of learning known in other animals. Yet it seems that there is much in the records that is of theoretical interest for comparative psychology and that a full discussion of the problems involved is warranted. My own interest in the subject is the outcome of work on the olfactory responses of the ichneumonid *Nemeritis canescens* (Grav.) and the dipterous fly *Drosophila melanogaster* Mg. and their experimental modification. The former experiments yielded results which were obviously inexplicable on the basis of simple kineses, taxes or even of simple associative learning, and so my attention was directed to the whole problem of learning in the Arthropoda.

But before we can proceed there are a number of terms which should be defined at the outset. This requires more than usual care owing to the fact that many of the terms employed are, or should be, used by the comparative psychologist in a precise and somewhat different sense from the loose usage of ordinary speech. A great deal of what has been written on animal behaviour is vitiated by obscurity regarding the exact connotation of the expressions used. I hope that I may have avoided some at least of such pitfalls.

The first term which needs definition is of course the term *learning* itself. Very briefly, 'learning' as used in this article implies *that process within the organism which produces adaptive change in individual behaviour as a result of experience.*[1] It will be noted that 'motor learning', or 'behavioural inertia' as it is sometimes termed, many examples of which could be quoted from among the Arthropoda, is ruled out by definition, since the phenomenon is essentially non-adaptive, being in fact a hindrance to rapid adaptive response.

Habituation may be regarded as the simplest type of change in behaviour which can be regarded as evidence of learning. The word is here used solely in its technical sense implying a simple learning *not* to respond to various mild shock or 'warning' stimuli, such as a flickering shadow or mechanical shock, when these are repeated without 'reinforcement', to use Pavlov's term, by any unfavourable or harmful result. Thus the word does not imply habit formation in general but rather the reverse. Humphrey (1933)[2]

[1] For more extended definition see Koffka (1935) pp. 529–30.

[2] References will be given at the end of Part III of this paper.

has defined habituation as "an instance of the living system's power of establishing a conservative equilibrium to change of external conditions". But even this definition requires qualification, since as it stands it includes a variety of peripheral nervous phenomena which have been given various names such as 'accommodation', 'acclimatization', 'negative adaptation', 'sensory adaptation' and even 'fatigue',[1] phenomena which may be the expression of the same fundamental property of living matter but which are quite obviously brought about by means of vastly different mechanisms. For this reason the term 'habituation' should, as in this article, be restricted to those changes of a relatively enduring nature which are evidence of the action of the central as opposed to the peripheral nervous system. Even so, it is not contended that even all forms of true central habituation, e.g. that of a snail to a repeated tapping and of a human being to the chiming of a clock or the roar of traffic, are identical. Obviously they could hardly be so since the complexity of the mechanism must vary so greatly from group to group. We may therefore define *habituation* as *an activity of the central nervous system whereby innate responses to mild shock and warning stimuli wane as the stimuli continue for a long period without unfavourable results*. We should also note that habituation need only involve stimuli from one receptor field, whereas the simplest associative learning must be concerned with at least two sets of receptors.

Instinctive reactions are so characteristic of arthropod behaviour that we can hardly discuss learning intelligibly without first defining what we mean by 'instinct'. The concept of instinct has been the subject of recent discussion by a number of writers, particularly by Wallace Craig (1918), Bernard (1924), Lorenz (1935, 1937, 1939), von Holst (1937), Heinroth (1930, 1938), Bierens de Haan (1940) and Armstrong (1942). Perhaps the best concise definition that can be given is to say that *'instinct' is an inherited psycho-physical disposition*, a concrete fact of mental structure which we infer from facts of behaviour and experience (McDougall, 1936). There have been attempts in recent years (e.g. Warden, Jenkins and Warner, Vol. I), to dispense with the term altogether and to assume that the word involves no concept which cannot be more precisely expressed in other terms. The failure of these efforts has however been made clear on the physiological side by Lorenz (1939) and on the psychological side by McDougall, Bierens de Haan etc. But the short definition given above needs amplification. If we follow Lorenz we can define an instinct as an inborn movement form or automatism, usually complex and resting on endogenous stimulus production. Instincts thus differ from taxes which are orienting reactions guided by external stimuli. They differ from chain reflexes in that they do not obey the 'all or none' law. Thus the longer an instinct is not released, the greater the intensity with which it will occur in response to a given stimulus situation; and conversely, 'damming up' an instinct by withholding all appropriate releasing situations results in the accumulation of reaction specific energy and a steady lowering of the threshold, giving rise to appetitive or goal directed behaviour (drive), until eventually the automatism provides its own adequate (endogenous) stimulus and the pattern of behaviour forces itself out, so to speak *in vacuo*. Thus from the psychological side instincts appear to involve an inner tension or feeling resulting in attention to objects of a certain kind or in a certain setting and to a conative urge in connection therewith. The above definition of Lorenz does not of course imply

[1] Sherrington ((1911), pp. 222–3) has given a warning against the uncritical use of the term 'fatigue', and points out that the waning of a reflex under long-continued stimulation is incorrectly so called.

that instincts are endogenous in the sense of being unrelated to the environmental situation. Though the appropriate environmental pattern or situation does not guide the automatism, it releases it by means of a corresponding 'inborn pattern' of reflexes (the 'coat of reflexes' of von Holst, 1937), inserted between it and the conditions of the environment. This reflex coat is capable of individual modification by various types of learning, in birds particularly by the curious process known as 'imprinting', and so serves as a somewhat flexible adjustment between the genetically fixed automatism and the variable environment. It may, as in the vast majority of insects, be very limited, rounding out the contour of the instincts, like a mantle of snow the contour of a roof. But however limited the coat the result is that nearly all examples of complex stereotyped behaviour represent intricate interlacement of reflex and instinct. The use of the term *imprinting* is confined to § VIII and its definition is more conveniently left until that place.

The terms 'trial and error learning' and 'associative learning' are used in their well-understood and traditional meanings which should be clear from § IV; so no definition is required here. Examples of *latent learning* will also be found in § V. This somewhat unsatisfactory term may be defined as *learning without patent reward*. It is characteristic of animals which show discrete exploratory behaviour dissociated from any immediate primary need or drive. Latent learning experiments seem to involve the preliminary learning of a locality or environment as a whole together with the ability subsequently to *select* certain parts of that whole for use as guides or signals for the whole or for some other part thereof or in connexion with a new motivation. The term *selection* is used in this article only with this special meaning. This power of latent learning and selection may at first seem to be a case of insight learning, in that by its means the animal is enabled to give an immediate adaptive response to what appears to be a new situation. Actually it is to be noted that in the case of latent learning the situation is never really new but is merely a selected part of a previously learned whole.

The term *Einsicht* was originally popularized by Köhler (1921), and was translated *insight* in the first English edition (1925). It implies the grasping of the relevant or essential relations between the features or objects entering into a total situation (McDougall, 1930). Insight thus corresponds to the 'articulate correlation' of Hobhouse (1927), involving apprehension of relations, whereas all forms of associative learning would belong to a lower category of learning, namely, 'inarticulate correlation' or 'assimilation'. On this view, which is apparently also held by Koffka (1935, pp. 382–3, 645), there is thus a real qualitative difference between 'insight learning' and associative learning. But there is another group of psychologists represented by McDougall (1936), Humphrey (1933), Tolman (1932) and Adams (1931), who regard all learning as in some degree the manifestation of insight. The members of this school argue that the well-established usage of contrasting 'trial and error' learning and 'insight' learning is unsound; the only difference, in their view, between the two categories being the rate at which learning takes place. If it occurs slowly we say the animal learns by 'trial and error'; if rapidly we say it shows insight. Adams (1931) makes the point that 'trial and error' is merely a means by which the experimenter measures the rate of the learning process and must not be confused with the learning process itself. He argues that in what we call 'insight learning' the rate is sufficiently rapid to enable us to see the true nature of the learning process of which it is an extreme case. Adams says: "The familiar fact of insight alone is sufficient to

demonstrate once and for all that there is no necessary or causal relation between learning and repetition or frequency of repetition."

Now this controversy (Hartmann, 1931; Ogden, 1932) is undoubtedly a fundamental one and theoretically of great importance, but I feel that in practice it need not detain the experimental biologist unduly. It has arisen in part I believe through confusion between insight *as the result of* a learning process and insight *resulting in* a new response. In fact, the remarks of Russell with reference to foresight (p. 224 below) apply also to insight (see also Russell, 1938, p. 164). Whichever view we incline to we must I think admit that slow associative learning resulting in 'inarticulate correlation', that is, without insight coming in at all, either before or after, is at least theoretically possible. We may also admit that the development of insight *as a result of* a process of trial and error or association is conceivable. This may in fact be what is involved in the learning of a pattern as a whole (cf. Lashley, 1942) and which is evidenced by the 'selection' characteristic of the classical latent learning experiments. Short of successful performance in such experiments it is doubtful whether, e.g. in most invertebrates, we have any definite evidence that insight can actually be produced in this manner. All we can say is that some invertebrate trial and error learning (e.g. that of *Astacus* (Gilhousen, 1929, fig. 2)) results in a response which, while 'perfect', is very far from being stereotyped as to detail.

But besides this possibility of insight arising as a result of the slow trial and error learning process, without being such as to enable the animal to cope immediately with a new situation, there is also *insight learning* in which insight *precedes* the 'solution', enabling an animal to bring together into a new combination a number of different pieces of past experience resulting, without any trial and error behaviour, in a new adaptive response to a new situation. Therefore for present practical purposes we define the criterion for the existence of '*insight learning*' as *the sudden adaptive reorganization of experience, or production of a new adaptive response, in circumstances not previously encountered* (see also Ogden, 1932). This true 'insight learning' is of course what we see demonstrated so perfectly in the classical experiments of W. Köhler (1921) with apes, where there is no preliminary trial and error behaviour, but in which insight into a situation results in a *sudden* learning of the problem producing a real solution. It must be borne in mind that (as Köhler has pointed out) most types of maze and puzzle box that have been used to study animal learning cannot test insight learning though the technique can of course be modified to do so (e.g. Shepard's test of reasoning; see Maier & Schneirla, 1935, p. 468). Thus, while with a typical maze experiment the probability of 'insight' being involved is directly proportional to the steepness of the learning curve we can never be sure of the existence of true 'insight learning' unless the chance or accidental discovery of the complete 'solution' by the animal has been virtually excluded.

Lashley (in Kluver, 1942), in a very suggestive and penetrating discussion of the problem of visual generalisation and the stimulus equivalence of similar visual patterns, provides a good deal of evidence for the view that something essentially similar to insight is involved in all recognition of visual pattern and is therefore present at least in all animals capable of form sight. He concludes that this can only be interpreted physiologically in terms of a basic conception of nervous function involving something in the nature of interference patterns.

In connexion with these divergencies of view among psychologists, the simplicity of the perceptual world of many invertebrates must perpetually be borne in mind. It may

indeed be difficult to picture pure trial and error learning, without even a modicum of insight, as being sufficient to carry us very far in accounting for the learning of the mammals. In fact we may guess that an enormous amount of so-called trial and error learning of the higher vertebrates does in fact involve insight at some stage. But this does not mean that true trial and error learning is necessarily impossible. The simpler and more restricted the perceptual world of an animal, the more efficient does the simple associative principle become; and in animals of simpler organization without form sight and where the sense data are fewer or where the attention is instinctively concentrated on certain aspects of the environment only, the possible combinations are much fewer, and accordingly the trial and error principle can carry us much further.

Some psychologists (K. D. & W. McDougall, 1931) argue that the concept of insight must also, to be intelligible, include some element of 'foresight', which should be regarded as entering into the meaning of the former term. I feel fully the force of this argument, and it seems that this conclusion must follow from the arguments of Koffka and of Humphrey. But while insight should be held to include 'foresight' we may agree emphatically with the penetrating observation of E. S. Russell, who (1932) says: "both in insight and in trial and error learning foresight arises—*before* experience in the former case, through experience in the latter." Thus while insight must include foresight, foresight is conceivable without insight as an outcome of the lower grades of learning. Sherrington indeed (1911, p. 332) speaks of "instinct with its germ of anticipation however slight".

The McDougalls would accredit both insight and 'foresight' to all animals which build a nest of specific form. They argue that such nest construction "is only to be intelligibly interpreted as the expression of a conative urge guided at every step by foresight of the results to be achieved and by insight into the materials dealt with at each step". But the remarks on p. 222 above, pointing out the difficulty of proving in a given case that insight results from the trial and error process, obviously applies also to the activity of nest construction, where, by the manipulation of materials, there is ample opportunity for trial and error learning. But some power of generalisation there must be; and to the necessity for some degree of conation and foresight in the construction of certain complex nest patterns I would readily agree *without of course claiming any foresight of the biological end of the actions, namely, the hatching of the egg or the development of the offspring.*

Till recently all attempts to dispense entirely with insight or similar concepts in theories of vertebrate learning have been shipwrecked on the problem of visualisation or sensory transfer. But even if it should ultimately prove possible to express all types of learning and generalisation in terms of some single basic attribute of organized nervous tissue, it is quite clear that some classification will be required, separating and distinguishing different stages in the almost infinite range of complexity which the study of learning reveals. No apology is needed therefore for attempting to define exactly the limits of the various categories of learning.

To conclude this matter of definition of terms it may be as well to point out that the terms 'intelligence' and 'reasoning' have as far as possible been avoided in this article. 'Intelligence' is used in such a vague manner as to render it practically meaningless; it may, for instance, in some writers' usage mean merely 'the capacity to learn', while with others it may connote the highest type of reasoning. Since, however, the English noun 'insight' has no corresponding adjective the word 'intelligent' is sometimes necessary.

The term 'reasoning' is held to imply those higher human thought processes for which Hobhouse uses the terms 'correlation of universals' and 'correlations of governing principles'; categories which of course do not concern us here.

II. Habituation

Habituation has already been defined and for a detailed account of a typical case Humphrey's (1930) paper on the snail, *Helix albolabris*, should be read. Many examples from widely separated groups are also listed in Humphrey's book (1933) and Piéron (1920) discusses others. As the former author has emphasized, habituation is ubiquitous among animals with a central nervous system, and something very similar has been observed in organisms of lower evolutionary grades from the Protozoa upwards, though here obviously the mechanism must be very different.

Some clear and rather instructive examples of habituation are furnished by my own experiments on *Nemeritis* and *Drosophila*. It was shown, for instance, with *Nemeritis canescens* (Thorpe, 1938) that the natural repellent effect of the odour of cedar-wood oil could be transformed into a high degree of tolerance, amounting almost to indifference, simply by exposing the adults, shortly after emergence, to an air stream bearing the odour of the oil (Fig. 1, Curve 3). It is clear that in describing such a result one need not assume any type of learning higher than habituation. In the case of *Drosophila melanogaster* (Thorpe, 1939) it was shown that not only could the repellent effect of an odour (in this case peppermint) be reduced but that it could be entirely eliminated and replaced by a significant positive chemotaxis. In this case also the effect could be produced either by impregnating the larval food medium with the peppermint or alternatively by exposing the flies, shortly after emergence, to an air stream bearing the odour. Now if it had been possible to find a pure 'artificial' odour to which the animal was initially indifferent over all concentrations, but which could, by the experimental situation, be converted into an 'attractant' we should clearly have a situation involving more than habituation. But so far, the search for such a substance has proved fruitless. It appears in fact that the most likely explanation of the *Drosophila* results is to be found in the fact that the peppermint used contained small quantities of ester as well as menthol. This latter is a repellent and is probably alone responsible for the repellent effect manifest on first exposure. The ester, on the other hand, is attractive but its attractive effect is at first completely masked by the presence of the menthol. But continued exposure, by analogy with the *Nemeritis*—cedar-wood oil experiments would almost certainly result in habituation to the repellent component with the consequence that the attractive component will now be free to manifest its effects, and we thus get the *appearance* of a complete reversal from repulsion to attraction, whereas in fact habituation alone is responsible. Recently Cushings's (1941*a*) results comparing the oviposition behaviour of *Drosophila* reared on a medium containing fungi, with those reared on an artificial medium, have given valuable confirmation of my own conclusions on pre-imaginal olfactory conditioning in this insect; here too it seems possible fully to explain the observed effects on grounds of habituation alone, assuming that the artificial medium employed gives rise to a mixed odour comprising both attractive and repellent components. Other results with *Nemeritis* (Fig. 1, Curves 1 and 2), to be dealt with in Part II of this paper, do however involve a higher grade of learning.

Humphrey, Piéron and others give examples, drawn from various animal groups, of

habituation to visual, auditory, thermal and tactile warning stimuli, the latter group including cases of adaptation to undeniably harmful and even fatal stimuli. While it appears that before the *Nemeritis* experiments no instances involving habituation to olfactory stimuli have been specifically put on record they must obviously occur. For instance, it seems that 'habituation' must often be involved in the experiments of workers in animal food preferences such as accustoming insects to new food under experimental conditions (e.g. feeding *Carausius* on ivy (Sladden, 1934, 1935)).

Innumerable further examples of probable habituation in insects could, if necessary, be culled from the literature, but it will suffice to mention a few. Harper (1907) has described a very clear case of the habituation of the aquatic larva of *Corethra* (Diptera) to mechanical vibration, and auditory habituation is known in the cricket, *Lyogrillus campestris* (Baier, 1930). Similarly Turner (1913, p. 350) implies habituation in the cockroach as evinced by the process of becoming 'tame'. The work of Wigglesworth (1941)

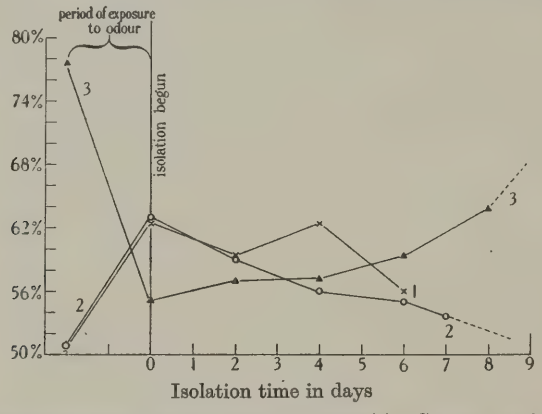

Fig. 1. Curves showing habituation and latent learning in *Nemeritis*. Curves 1 and 2 show latent learning by *Nemeritis* resulting from exposure to the odour of *Meliphora* larvae; results expressed as percentage attraction to the odour, before and after exposure: (1) period of preliminary exposure, 48 hr.; (2) period of preliminary exposure, 6 days. Curve 3 shows habituation of *Nemeritis* to odour of cedar-wood oil expressed as percentage *repelled* before and after exposure. (After Thorpe, 1938.)

on the human louse provides some excellent instances of the habituation process. Among the Arachnida the enduring habituation of the spider, *Cyclosa conica* Menge, to the sound of a tuning fork was established by the Peckhams as long ago as 1887.

The question how far the process of habituation can be regarded as equivalent to the "experimental extinction" of Pavlov is a difficult one and is somewhat outside the immediate scope of this paper. Pavlov (1927) and Beritoff (1927) claim that the extinction of the orientation response in the dog, in other words habituation, is identical with the extinction of the conditioned response. But, as Humphrey has pointed out, it seems clear from Wendt's (1930) monograph on the conditioned knee jerk that 'habituation' and 'experimental extinction' are not precisely the same, although it would be unjustifiable to assume that the mechanism is fundamentally different.

III. Habituation in ants and other social Hymenoptera

There is much apparently complex behaviour among ants that may in fact involve no more than habituation. It is well known that, as recorded by Fielde and a great many other observers, many species of ants exhibit antagonism to ants of other species and

even of other colonies of the same species. It is clear from the work of Bethe, Fielde (1904) and others that this antagonism is mainly if not entirely a question of distinctive odour. According to Fielde (1901) the antagonism of workers of an ant such as *Stenamma fulvum piceum* is aroused by any ant odour which a given worker has not previously encountered individually. But it is also well known that under various conditions artificial mixed colonies of ants can be established, and that once the individuals of the two species or races or nests, as the case may be, have become thoroughly accustomed to one another the new composite nest is often a stable and permanent entity. Brun (1912) has shown that the circumstances under which there is a good chance of an artificially mixed population of ants becoming rapidly 'allied' are as follows: (1) when both parties are brought into an equally difficult or unusual situation; (2) when both groups are in approximately the same numerical strength and when one group at least is in possession of numerous brood; (3) in the presence of fertile queens; and (4) in the presence of a powerful common enemy. Now it is clear from Brun's experiments that these results cannot be due to a simple associative olfactory conditioning between the individuals of the two groups. That this cannot be the explanation is also suggested by some as yet incomplete and unpublished experiments of my own on *Acanthomyops flavus*. If then typical 'associative learning' will not explain such observations, on what basis can they be interpreted? Miss Fielde's explanation involves very elaborate psychological concepts. She says: "Any distinctive odour to which an ant is accustomed and with which it associates security and satisfaction is attractive to it, while ant odours to which it is unaccustomed excite alarm and hostility in proportion to their strangeness." This theory, though conceivably true of ants in some instances, clearly does not apply to the conditions which obtain in the four situations instanced by Brun (above). These ants are certainly not experiencing "security and satisfaction". No; what is to be noted about all these sets of conditions under which permanent mixed colonies can become established is that there is in every case some distracting overriding feature which by its greater urgency inhibits the normal pugnacity for a period, just as the beekeeper when mixing stocks of bees distracts them by sprinkling them with dust so that the cleaning reflexes for a time dominate all else. During this period the process of habituation is at work and quickly dulls the strangeness of the new odour so that it becomes accepted ('habituated') as the normal background of the new environment and so no longer gives rise to animosity. It seems highly probable that 'habituation' will also account for the instances of acquired tolerance in other hymenopterous groups. P. Rau (1928) records much the same phenomena as between rival queen *Polistes* wasps, and Stöckert (1923) says that rival sister queens of *Halictus* "*may compose their differences*" and establish a colony on the basis of dual monarchy. Finally, to return to ants once more, it is noteworthy that Fielde (1901) comments upon the increasing 'tameness' of ants which are frequently handled, and Donisthorpe (see Hingston, 1928, p. 187) tells how a colony of *Myrmecina graminicola* kept in confinement for many years "lost their natural timidity".

IV. Associative learning

Under the term 'associative learning' are grouped a number of clearly different phenomena of learning which it is not always easy to define sharply. These examples may vary from the simplest 'conditioning' to the more elaborate forms of 'trial and error' learning and the learning of simple mazes. That associative learning of one type or another is widespread

among insects is obvious. Clearly any learning process that is not, like habituation, of a negative character must at least involve simple association. But it will be instructive to consider exactly what the recorded examples of associative learning among insects and other arthropods do involve.

(1) *Conditioning*

First let us consider the true conditioned reflex. An elaborate discussion of the thorny problem of the relation of the classical conditioned reflex mechanism of Pavlov to learning and habit formation in animals would be out of place here. For a very profound examination of this matter the reader is referred to Chapter IX of Humphrey's book (1933). Suffice it to say that, following Humphrey, the conditioned reflex is for our purposes looked upon as "the ultimate word in simple associational learning. In it, simplification of the learning process has gone so far that the 'habit' of everyday life has lost some of its characteristics while retaining others....Habit as we know it outside the laboratory is certainly a much more stable affair than the evanescent conditioned response which, when repeated, progressively and spontaneously disappears. Probably it is best to think of the conditioned reflex as an *incompletely recorded* case of habit." It is a minute and artificially isolated piece of the learning process of the animal, and, while its study is of the greatest value for the better understanding of the learning process, it is not the whole learning process itself and therefore we must not expect to find that even the simplest learned responses of the whole animal under natural conditions can be equated with a simple conditioned reflex and nothing more. The claim that upon the conditioned reflex there can be founded a scientifically determinate psychology parallel to an exact physiology appears now to be illusory. This of course must not be taken to imply that a full physiological description of the higher types of animal behaviour is therefore impossible; for stimulating excursions along newer lines see J. Z. Young (1938) and Lashley (in Kluver, 1942).

Perhaps the simplest examples of associative learning in insects are those described in moths of the genera *Saturnia* and *Catocala*, by Turner (1914), who conditioned shock reactions to sound; in mayfly nymphs, *Heptagenia interpunctata*, by Wodsedalek (1912), who similarly conditioned the animals to a shadow stimulus; and in cockroaches by Szymanski (1912). Very similar experimental results have been reported by Mikhailoff (1922, 1923) and Cate-Kazejewa (1934) working with hermit crabs of the genus *Pagurus*. These cases are certainly in essentials very suggestive of the classical conditioned reflex mechanism even though vastly more stable.

(2) *Trial and error learning*

(*a*) *Simpler types.* From 'conditioning' we pass to 'trial and error learning'. Now there are, as Hull (1930) and others have pointed out, several different degrees of complexity in learning grouped together under the general term 'trial and error', which is usually held to include the mastery of quite complex mazes or elaborate puzzle boxes where the reward is obtained only after completion of a series or particular combination of trial acts. But even the very simplest of trial and error learning differs in certain important respects from conditioning, as Grindley (1932; 1937) has shown. He says: "A conditioned response may be thought of merely as the connection of a stimulus with a new response when the two are made to accompany each other in time. In trial and error learning on the other hand we have to think of the reward as exercising a *retroactive*

influence causing a stimulus to become connected and remain connected to a response when both the stimulus and the response *preceded* the reward in time and when the response was not originally a response to the reward" (Table 1). The essential difference in time relation of the two processes is well brought out on inspection of the time curve for conditioned reflex formation given by Hull in Murchison's *Handbook* (1934). Hilgard (1939) quotes Tolman and Schlosberg as supporting this view; but he himself, while agreeing that the two processes are not identical, comes to the conclusion that no sharp boundary can be drawn, and that it is important that more work be done on transitional situations. In this connexion work such as that of Brogden may prove of great value. Clearly the differences between the two are in part artificial, in so far as they are directly imposed by the limitations of the Conditioned Reflex experimental technique.

The simplest type of trial and error learning is that involved in straightforward food presentation tests. A good example of this is provided by the experiments of Mayer & Soule (1906) with the caterpillars of *Danais plexippus* which were able to develop an associative memory of a distasteful leaf (e.g. *Ampelopsis*) lasting for $1\frac{1}{2}$ min. In such an experiment we must suppose that the larva on being presented with the strange leaf perceives it, probably by its olfactory sense, as something to be eaten. It takes a bite

Table 1. *The difference between the establishment of a Conditioned Reflex and the process of Trial and Error Learning as seen in Grindley's experiments*

	Initial state	Training process	Result
Conditioned reflex	(Rwd.→Rsp.)	St.→(Rwd.→Rsp.)	(St.→Rsp.)
Trial and error learning	St., Rsp., Rwd. (all independent)	St.→Rsp.→Rwd.	(St.→Rsp.)

Note that in Trial and Error Learning the 'reward' never comes before the 'response'.
St. = stimulus (e.g. buzzer).
Rwd. = reward or reinforcement (e.g. food, shock, etc.).
Rsp. = response (e.g. salivation, turning of the head, etc.).

with unpleasant results. After a number of trials it comes to associate some specific odour or other quality of that particular kind of leaf with the unpleasant effects experienced subsequently upon eating it, and is thus enabled for a short time to avoid it. Thus the 'reward' or 'reinforcement' here is the unpleasant taste and, as distinct from the course of events in a true conditioned reflex experiment, the reinforcement exerts a retroactive influence linking the new stimulus (specific odour) with the 'response', namely, mastication. Exactly similar experiments have been performed upon *Dytiscus* by Schaller (1926), the beetles being trained to avoid meat made bitter by quinine, and by O. Koehler (1924) on dragon-fly larvae (*Aeschna cyanea*). All these are instances of negative food presentation training, i.e. training in which the incentive is avoidance of 'punishment'. It is interesting that Koehler was unsuccessful in conditioning his dragon-fly larvae to respond to a given colour without this punishment; the simple hunger drive (positive food presentation training) being quite ineffective.

(*b*) *More complex types.* These examples above quoted will serve to illustrate the simpler type of trial and error learning in insects and their relatives. The more complex types of trial and error learning differ from the above chiefly in the degree of complexity of, and the length of time taken over, the trial acts. The experiments of von Frisch (1914, 1931, 1937) and others on the honey bee, around which an extensive literature has grown, provide examples of the converse of the experiments just described; namely positive

food presentation training. As will be clear to those well acquainted with the training of bees, the orientation of these animals to the food place (just as to the hive) is in fact a highly complex matter probably involving 'latent learning' and 'selection' (see Part II of this paper). Nevertheless, isolated phases of the training process show behaviour which appears closely similar to trial and error learning and as such is appropriately recalled here. In these experiments the bees are enticed to find food (sugar water) placed in glass dishes on a certain table, and eventually come regularly to feed at the table. Sugar water is then provided only on squares of a certain colour, and immediately this colour becomes associated with the feeding orientation and all the other squares are ignored. Here, it is true, the retroactive action does not at first sight appear quite so clear since the colour is presumably visible not merely during the period of orientation but also during the whole time of feeding. Nevertheless, that it is a true retroactive action has been established by the work of Opfinger (1931) who showed that during the period of feeding the bee is uninfluenced by other sense stimuli. Schaller's (1926) success in training the water-beetle, *Dytiscus*, to come out of the water to feed on a certain platform of a projecting rock, parallels von Frisch's work very closely.

Maze learning is a characteristic example of this complex type where the *obvious* reinforcement of the conditioning process comes only at the end of a particular combination of trial acts. The very simplest types of maze such as dual compartment mazes and 'Y' and 'T' mazes, having only one choice point, would seem at first sight to offer problems little if at all more difficult than that involved in learning a simple food discrimination habit. But the length of time which elapses between the trial act and the reinforcement is probably the critical factor in the situation and, be that as it may, the fact that a dual compartment maze *is* more difficult to learn than a food discrimination habit is suggested by the work of Mayer & Soule on the larvae of *Danais plexippus* mentioned above. These authors found that the larvae failed to learn the direct path to food in a dual chamber maze, one chamber being barren while the other contained a growing food plant. Blees (1919) claims that daphnids can be trained to approach a light source by travelling part of the way through a glass tube, in such a manner that the number of collisions with the walls is steadily reduced with experience. Agar (1927), on the other hand, reports that water mites (hydrachnids), *Eylais* spp. and the Crustacea-Entomostraca, *Daphnia carinata* and *Simocephalus* sp., are unable to master the simplest Y maze even though varied sensory cues were supplied. These arthropods seem thus to be inferior in learning ability to earthworms which, as is well known, can learn a maze containing a single choice point. The Malacostraca, on the other hand, can master simple mazes easily and can be taught relatively complex mazes (see Gilhousen, 1929, *Astacus*; Agar, 1927, *Parachaeraps*; Cate-Kazejewa, 1934; and a number of earlier workers). Among the Myriapoda, Scharmer (1935) has shown that *Lithobius* can learn a T maze provided the two walls differ in texture. This result should serve as a warning against accepting at their face value experiments which merely report failure of arthropods to learn mazes. It may often be that the experimenters have failed to provide the animal with the type of cue which it is able to use, or have not provided the right motivation. This question of appropriate motivation requires much more careful consideration on the part of workers on arthropod behaviour than it has hitherto received. Thus the failure of Mayer & Soule (above, p. 229) to train the larvae of *Danais plexippus* to take the direct path to food may have been exactly parallel to Koehler's failure in positive

food presentation tests with *Aeschna* larvae and may have had nothing whatever to do with the kind of 'maze' used. It may be that with the less highly organized arthropods avoidance of 'punishment' is a more satisfactory incentive than is any kind of 'reward'. But much more work is necessary before such a conclusion can be considered as established; and, as P. T. Young (1936) has pointed out, the distinction between punishment and reward, while practically important, is not psychologically fundamental in experiments with animals.

Among adult insects it is known from the work of Turner (1913) and of Eldering (1919) that cockroaches can learn relatively complex mazes; and ants, as the work of Schneirla (1929, 1933, 1933*a*) has shown, can master a maze containing up to ten 'blinds'.

It has already been shown (p. 226 above) that habituation will go a long way towards explaining many of the learned responses of ant colonies, but this is not to deny that trial and error learning must play a large part in ant life just as it does among other social Hymenoptera. Indeed, it is doubtful whether the existence and maintenance of social life of any complex kind could be envisaged without postulating some capacity for association in addition to instinct and habituation. The ubiquitous phenomenon of trophallaxis in fact provides almost unlimited opportunity for the establishment of associations through simple trial and error learning. But the behaviour of the more highly developed types of ant colonies may in fact involve much more and, as we shall see below, there are some indications that 'latent learning' may well play its part in the organization of such societies. There are no valid grounds, however, for accepting Verlaine's (1932–4) conclusion that the organization of aculeate societies is maintained by transmission to "descendants by example and imitation." Nor does there seem any sufficient evidence for adopting the strongly anthropomorphic position of Lafleur (1940, 1941).

How far simple trial and error learning will suffice for the mastery of the more complex mazes is a difficult problem. The chief drawback to the maze method is the difficulty of analysing the process involved in the complex response of running the whole labyrinth. Nevertheless, the technique has one substantial advantage; namely, that the problem set is one that for many animals is a very natural one. Many animals have to find their way about in nature, and to ask them to learn a maze is to put them a problem in solving which they are likely to be able to show their capabilities to the full.

Most workers seem agreed that whether or not a complex maze could be learnt by a series of separate and discrete associations for each choice point, the behaviour of vertebrate animals such as rats and mice in the maze does often indicate more than this; and that it is necessary to assume that by the time the training is complete we have integration of experience resulting in response to the pattern of the maze as a whole. But though this may be true of mammals and is suggested by the results with ants described below, there is nothing which conclusively proves this in the meagre records of maze running among other insects or arthropods, though the interesting observations of Gilhousen (1929) on the maze learning of the crayfish, *Astacus trowbridgii*, must be borne in mind here (see Fig. 2 and p. 230 above).

The only insects which have been taught to run complex mazes are the cockroaches, *Periplaneta orientalis* L. (Turner, 1913) and *P. Americana* (Eldering, 1919) and the ants, *Formica subsericea* and *F. pallidefulva schaufussi* var. *incerta* (*F. inerta*). The maze used by Turner in his training of *Periplaneta orientalis* is shown in Fig. 3. Eldering's results

are indicated by Figs. 4 and 5. Both Turner and Eldering comment on the great individual variation in learning ability displayed, a variation that seems to have no relation to age; and similar variation in ant learning ability is described by Schneirla. While various senses may be used, the sense of touch seems by far the most important in the maze learning of the cockroach, and care was taken to eliminate the possible effect of odour trails. Eldering shows that the habit is not destroyed by narcosis. Two examples of Turner's learning curves are shown in Fig. 6.

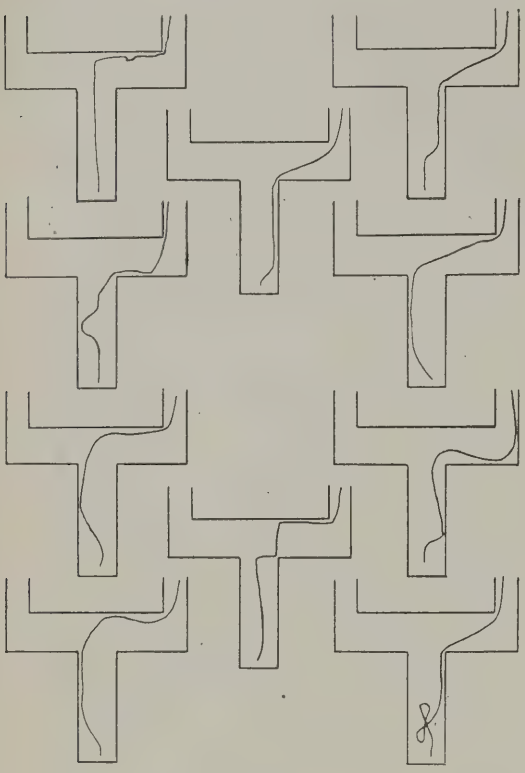

Fig. 2.

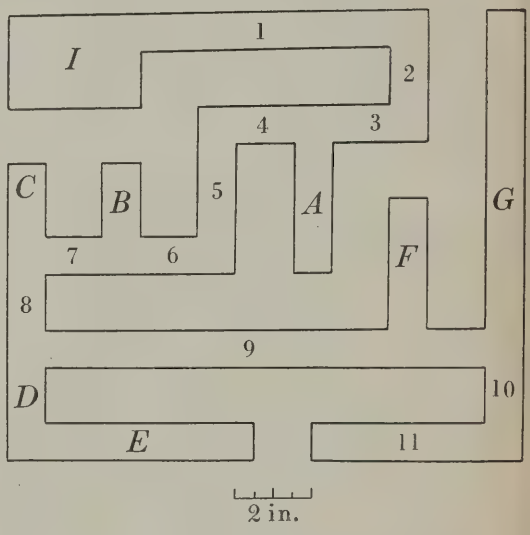

Fig. 3.

Fig. 2. Diagram to show variable behaviour displayed by a fully trained crayfish in a series of correct runs through a simple maze (after Gilhousen, 1929).

Fig. 3. Maze used by Turner in training cockroaches. *I*, platform on which animal is placed. *C*, goal (i.e. point of entry of home cage). The numerals 1–7 indicate the direct passage way to *C*. Other letters and numerals denote blind alleys.

That learning capacity of a fairly high order must be involved in the orientation of ants has long been evident from the observations of many workers who have studied the insects foraging under natural conditions. The well-established ability of the individual worker to straighten her path with experience, to avoid long detours in favour of a more efficient path from one point to another, and to eliminate excess back tracking, all point to the conclusion that ants should prove quick and capable at learning mazes, and the pioneer investigations of Turner (1907*a*) and of Schneirla (1929, 1933, 1933*a*), have shown that this is indeed the case. The extreme of maze complexity which can be mastered by exceptionally capable individuals of *Formica incerta* is shown in Fig. 7. The performance

of a 'fairly successful' individual of *F. incerta* in this maze is shown by the learning curve in Fig. 8. That the light compass method of orientation is of prime importance in the direction finding of ants as well as other homing Hymenoptera has been clear since the 'Fixierversuch' of Brun in 1914; and the fact that ants follow 'odour trails' has been known since the work of Bonnet (1779–83). But the great merit of Turner's and Schneirla's experiments is that they make clear the relation between visual, olfactory and other

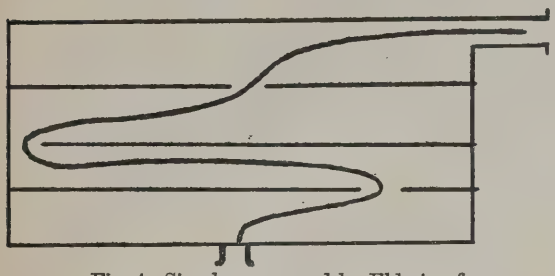

Fig. 4. Simple maze used by Eldering for training *Periplaneta*.

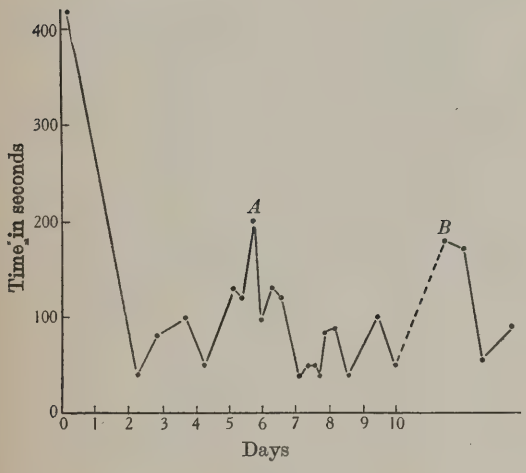

Fig. 5

Fig. 6

Fig. 5. Learning curve of *Periplaneta* (after Eldering), (mean of 5 animals): *A* = box turned through 90°; *B* = after lapse of 1 month without training.

Fig. 6. Learning curve of cockroach (after Turner): *A* = total errors; *B* = time in minutes required to run maze. Each curve represents average of 10 cockroaches.

clues in ant learning. Thus the direction of the light rays is effective for orientation whenever the differential aspect is marked enough; but when this is not the case, or when visual control is made unstable, other factors become more important. Thus the more often the direction of illumination is changed, the less reliable it becomes as an orienting factor, and the more the unchanged features are relied upon, and other senses, e.g. olfactory, tactile and kinaesthetic are brought into play. There is prompt recovery when original conditions of illumination are restored.

The maze shown in Fig. 7 is so designed as to provide both direct and alternative

routes on both the outward and return journeys of foraging ants and the right-hand half of the maze is a replica of the left. The ant can thus be set an identical maze problem either (1) on the outward or on the return journey alone (separate situation learning), or (2) on both together (combination problem), or (3) one after the other (successive presentation problem). It is found that the ant learns the pattern much more readily when returning to the nest carrying food than it does if faced with it on the outward

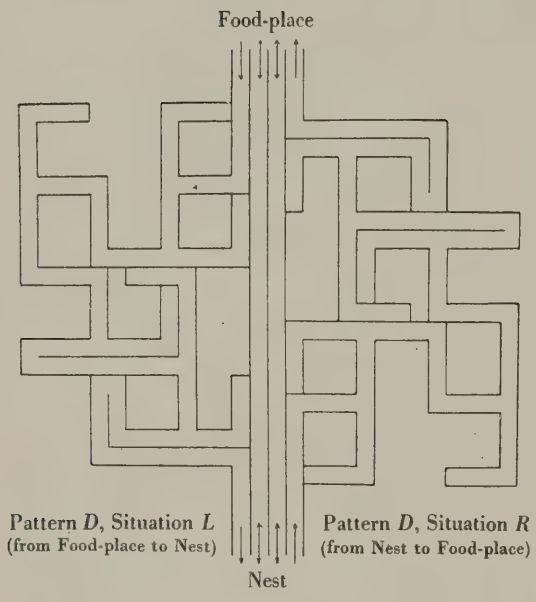

Fig. 7.

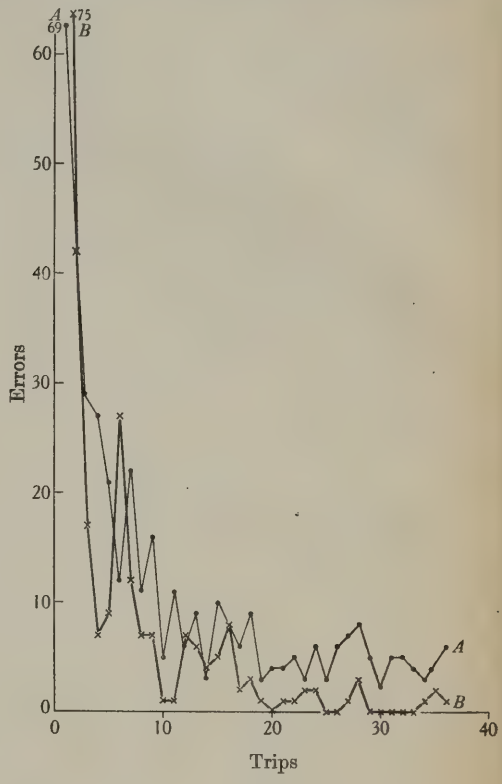

Fig. 8.

Fig. 7. Maze used by Schneirla (1933) in the investigation of ant learning.

Fig. 8. Learning curve for a fairly successful performance of *Formica incerta* in a complex maze (after Schneirla): A = outward trip to food place; B = return trip to nest.

journey. The complete combination problem ($DR+DL$, Fig. 7) proved too difficult for the ants to solve fully though one individual ant made very remarkable progress in it. In this relatively complex situation there is no evidence for transfer of training, for the problem of the return route is attacked essentially as though it were a new one. As will be seen later, this evidence of the independence of orientation on the outward and on the return journey receives some support from the work of Opfinger (1931) and Hannes (1930) on the forage flights of bees, and is best considered in connexion with the problem of homing.

NOTICE TO CONTRIBUTORS

Because of the restrictions on the amount of paper issued to the publisher, it will in general be impossible in future to supply the authors of papers published in this journal with any offprints additional to the twenty-five free copies. In special cases the author may himself apply to Mr W. Lewis, Cambridge University Press, who will supply additional offprints if he has any paper available; but Mr Lewis cannot guarantee to be able to supply such offprints, unless the author is able to provide the paper himself.

ERRATA

In the paper by D. N. Lawley, 'The application of the maximum likelihood method to factor analysis' in vol. XXXIII, part 3, of this *Journal*, the following misstatements occurred:

P. 173, line above Table 4	'rows (*b*) and (*c*)' *should read* 'rows (*a*) and (*c*)'
P. 174, line 15	'(*f*) and (*g*)' *should read* '(*e*) and (*g*)'

CRITICAL NOTICE

By F. C. BARTLETT

The Psychology of Early Childhood. A Study of Mental Development in the First Years of Life. By C. W. VALENTINE. London: Methuen and Co. 1942. Pp. xvi+557. 22*s.* 6*d.* net.

Probably everybody has had the experience of waiting for something to happen for a very long time, of getting keyed up with expectation, and then of suffering a collapse into disappointment when the long-delayed event is at length realized. The first part of this experience has, inevitably, been illustrated in the case of Prof. Valentine's book. Not only has it occupied twenty years of unremitting research, and of reading, study and serious thinking, but from time to time bits of it have taken a public shape, in this *Journal* and in others, and in discussions at a variety of centres of psychological teaching, whetting the appetite for more. But the second part of the experience, the disappointment, nobody who reads the book with care and with a proper regard for facts and their interpretations is likely to suffer. At first, perhaps, one regrets that the book has come when most of its potential readers all over the world have been so swept into matters of a very different kind that they can give it a far less leisurely and discriminating study now than they would wish to do. Even this, however, has its compensations. There is something refreshing, in these days, in finding temporary release from the demand for the short view, from the need to give swift answers to pressing problems the very character of which may have changed before solutions are available, and in returning for a time to the calm atmosphere of genuine scientific research.

It may be that no book is easy to review. This book certainly is not. Choice must be made between two possible methods. On the one hand, a digest of the whole of the contents of the volume may be attempted; or, on the other, drastic selection may be applied, and particular issues and problems discussed in the light of Prof. Valentine's treatment of them. To do both in one review is next to impossible, for very nearly every major problem in the psychology of childhood is discussed with abundant illustration.

I have deliberately adopted the second method, largely because I believe it to be the one more in line with Prof. Valentine's own approach. He is obviously interested in the reactions of his own five children and of the very many others who, directly and indirectly, have contributed to his studies, not only for their own sake, but also because of the light which they throw upon unfolding development and the understanding of human behaviour generally. Sometimes, indeed, as in the exceedingly interesting chapter on Laughter, it seems as if the adult expression comes right to the front, the factual materials being regarded chiefly as forward-looking incidents, foreshadowing the functions of what is treated as the same behaviour in the more mature community. To my mind it is this willingness to search for the full significance of the data dealt with in an ordered movement of growth which particularly gives the book its high rank among studies of childhood, and establishes an obligation upon psychologists of all types, and not only upon students of child behaviour, to read and re-read it with critical care.

One fact becomes apparent in the first few pages of the volume and is exemplified again and again in every chapter. This is the diversified, complicated and organized endowment which even the youngest child brings with him to meet the demands of his environment. I doubt whether Prof. Valentine's general scheme quite does justice to his facts in this connexion. He is more tied than he need be to the conventional psychological pattern of an order of itemized reactions, of reflexes, of instincts, of defined emotions and, as they come along, of categorized cognitive behaviour. He has dealt, I hope, a final death-blow to the dogma that to the infant the world is just a "big, booming, buzzing confusion". The incidents of infant behaviour, when they flash out, are very often as definite, and serve as exact an end as any other type of behaviour. But the reaction appears, disappears, reappears not as if it expresses some underlying and rather fixed system of basic movements, fundamental situation responses and ways of thinking. Thus while there is little in the picture to suggest anything like an 'undifferentiated continuum' to begin with, slowly changing into highly specific responses by the exercise of some single activity of discrimination, there is also not much, either, to suggest a few reactions of

fixed but general character as a basis, developing by some internal process of 'maturation' into a behaviour which is at once extremely varied and remarkably uniform.

The basic endowment out of which behaviour grows appears far more likely to be of a kind recently suggested by Dr C. S. Myers[1]: an inherited "arrangement" of sensory, ideational, conceptual "and even of conative 'material'". The key process throughout seems to be adaptation, present to begin with, but developing, with many twists and turns, spurts, stops and restarts, always according to the same principles, though with surprising individual flexibility, until the widest range of adaptation of which the psycho-physical organism is capable having been reached, fixity and its accompanying degeneration set in. All this does not mean that the notions of reflexes, instincts and many other varieties of categorized reactions must be abandoned, but that they need a radical revision in the light of their functional significance.

There is astonishingly little in this book about temperament, and nothing at all about character. It would no doubt be unfair to infer that Prof. Valentine considers that children up to the age of five have "no character at all", or that temperament plays little part in determining how they behave. Actually what is said about temperament is singularly interesting. Eighteen temperamental qualities were listed and five children ranked, or rated, for their possession of them. The qualities were found to be common to all the infants, but their arrangement and relative weighting differed in each case as compared with the others. Little or no attempt is made to follow up this clue, or to discover whether the appearance and disappearance of later specific reactions and their particular characters were constantly related to this temperamental setting. For example, in his very important chapter on The Innate Basis of Fears, Prof. Valentine rightly pays more attention than anybody else has yet done to the fact that some children seek fear. "Several children were observed to enjoy 'playing at fear', even to an extent which eventually caused screams and tears." His chief comment is that "this suggests an innate craving for stimulation". Why should it not suggest something neither as itemized nor as simple as this? Not many cases are given, but they all seem to me to suggest a temperamental *arrangement* which, facing the challenge of an outer world, develops the adventurer, and at the same time determines the special directions in which he will "play at fear".

The long chapter on Reflexes is full of first-hand observation, of excellent experiment and of judicious selection from a wide range of records. It shows how unstable these small items of conduct are in the case of the very young child. Their date of appearance varies considerably in different cases. They may be exhibited, then die away and then apparently return, but often this time with a rather different setting, as if the underlying functional tendencies have become broadened. They may be expressed partially only—this is extremely well worked out in the case of the eye-blink. Their inhibition and its probable relationship to central processes are admirably discussed; less fully, the connexion of their reappearance with the growth of the central nervous system. The fact that they have a vital setting, of affective or cognitive process, or of some directed stream of activity more general than themselves, is beautifully illustrated. All this material is used to lead up to a vigorous and merited attack upon the sharp distinctions which have often been made between reflexes and related forms of behaviour such as the instincts.

In the end, however, Prof. Valentine seems to want to keep the old distinctions in the old way, and only to abolish some of their less reliable defining characters. He says that reflexes show no persistency with varied effort, or none that results in "adaptation of the response to varied situations". So far as I can see his cases do in fact often show both consistency and variability and both together. If he had measured his reflexes more exactly he would almost certainly have come across more variability still. And if reflexes are not persistent why do they often drop out and then come back again, or must we suppose that their return is nothing but a "first appearance" of something else? Moreover, reflexes are said to lack "self-originated seeking for the stimulus". This is a slippery criterion, and I cannot find that Prof. Valentine anywhere shows that it is an unambiguous determinant of any form of human action whatsoever. Not a little of the behaviour characteristic of "playing at fear" is on what most people would call a reflex level, and it looks as if this is as near a "self-originated seeking for a stimulus" as one might expect to find. If it were said that here the reflexes have their "instinctive" setting, the reply would be that every reflex has been shown clearly to possess its wider functional background.

Let me add that the argument against treating the whole of behaviour as a growing complication of conditioned reflexes is, to my mind, triumphant.

[1] *Brit. J. Educ. Psychol.* (1942), XII, 155.

When I had finished this chapter, for which I have a great amount of admiration, it still seemed to me that Prof. Valentine had all the material available to effect a considerable reshaping of traditional views, but for one reason or another shrank from attempting this task.

I should like to try to write something about the chapters dealing with Play, Imitation, Fears (this one especially), Laughter, Suggestion, and various forms of special emotion. Also I should like to comment on the detailed criticisms of the Freudian point of view, most of which seem to me to be just, though many of them must inevitably turn upon the interpretation of ambiguous evidence. But I propose instead to pass on to the discussions of what are often called the "higher mental processes".

The chapter on Learning and Remembering is rather slight, and on its face is very largely concerned with dates. Yet all the illustrative material is full of interest. It shows that "association, recognition and recall" begin very early and are most highly selective from the outset. It demonstrates that while repetition, and especially exact repetition, is less effective than many people have thought, there does appear to be an early age during which repetition—together with something else—is required for any form of lasting effect. In this discussion, as in many others, it is indicated that something happens during the 3–6-year period which immensely facilitates all these processes, though whether the essential factors in this development are mainly individual or mainly social is not settled. The cases, perhaps, suggest the latter.

The study of Ideational and Thought Processes, together with the following and closely related consideration of Some Special Abilities and Ideas, though the order of their arrangement pursues a somewhat conventional course, I found of fascinating interest. Once more Prof. Valentine is considerably preoccupied with dating the emergence of different sorts of meaning, images, ideas, relational references, generalizations, inferences and types of reasoning. The dating is, however, used as a basis for many illuminating comments about the developing life of the child in its dealings with the outer world. In this Prof. Valentine seems to me to be vastly more successful than Piaget, who is kindly, but firmly and rightly, criticized at many points. The method of investigation is a blend of observation and experiment, and many of the experiments are original and offer a field for further exploration.

In all this discussion of the higher mental processes the point which I have mentioned already stands out afresh. Every distinction, upon a variety of uses of which the higher developments of mental processes depend, is foreshadowed in expression or behaviour, and is present in a world of perception before, in a new form, it is used in a world of ideas, or of reasoning. Moreover, the illustrations show decisively how these distinctions are not—except, perhaps, for the unsympathetic or concept-ridden onlooker—present in a woolly or confused manner. They serve functions just as definite in the early stages as they do in the later. Formulation is achieved as the functions broaden, and through it the usages which express the functions spread out, and the range over which situations become the stimuli for adjusted and adaptable behaviour increases.

One important question which is raised by these observations is, what are the psychological status and uses of these early foreshadowings of subsequent behaviour? Prof. Valentine often writes of the sensory or perceptual expression as "involving" the later ideational one. This distinction, common as it is, has always seemed to me to be difficult to comprehend. At the age of a few months, one of Prof. Valentine's children "looked at me in a very puzzled, concentrated way, when he saw my face covered with soap lather for shaving". At a little over two years the same boy "in judging a circle said, 'O not a (s)quare'". Later still, no doubt, he arrived at the stage at which all circles are treated as different from all squares, and at the notion of difference as applicable to an infinite variety of cases. It is very hard to see how the last stage can be "involved" in the first in any but a purely formal sense. It looks as if, at each stage, the particular quality of reaction is admirably suited to the interplay between mind and environment as both of them develop, but also that at each stage its predominant quality is different from the preceding one.

The child is born ordinarily into a world which is regularized for him as completely as possible both by the care of adults and by his own relatively limited avenues of approach to what goes on outside. Feeding, sleeping, playing and the few other activities all have their times and seasons and come and go regularly. Prof. Valentine, with his singularly inquiring mind, jogged his own infants out of this regularity often enough, and in so doing probably pre-dated some of their discoveries and excursions, though in no undesirable manner. But the changes and accidents of the world will do this anyway, and when they do, out, for a moment, will flash the reactions which are the forerunners of future routine. When this occurs the expression will have its own character, telling, if it is considered in relation to the effective situation, its own

clear story of the phase of growth reached. There may be just puzzlement in the case of difference, or a fleeting shadow of a smile when something strange passes swiftly into an already appreciated pattern. Gradually the organic needs and the tendencies that cluster about them become more fully interrelated and individualized, and now they begin to make their own demands more insistently. As this proceeds puzzlement, for example, is not simply unrolled, but, without being superseded, it changes in some cases into the cue for comparisons and negations. There is no continuance here either. The growing organization of needs and tendencies which first seems as if it finds its satisfaction in splitting the world into this and that and resolving it into a mass of instances, carries with it its own dissatisfactions too. For round about it are lots of other organizations of needs and tendencies, similar, perhaps even identical in the items of their make-up, but each group characteristically arranged. And the one individual organization must solve the problem of living with many other individual organizations in a new social world. In the course of doing this it must discover the ways of generalization and of the appreciation and application of the universal relations.

At every stage cognitive achievement swings from routine to diversity and back to routine again, at the command of that background of needs and tendencies which set the course of the individual's life. Every stage is a satisfaction and a search, a finding and a disappointment, until the pattern of tendencies sets and, sooner for some, later for others, routine holds sway, only to be broken up again by the next succeeding generation.

The Psychology of Early Childhood, with all its wealth of evidence, shows that there are a remarkable order and dating in this process, side by side with an equally persistent and remarkable organization into more and more individualized expression of its determining tendencies. Looked at in this way, Prof. Valentine's all too brief Appendix on Stages of Development is full of significance. He has four main rubrics: (1) General and Social; (2) Innate Propensities, Feelings and Emotions; (3) Movements and Skill; (4) Learning, Remembering, Language and Ideas. During the first few months of human life the General and Social heading seems to demand the lion's share of the space, and even the entries under the other titles—a possibility which Prof. Valentine himself points out—seem many of them as if they might just as well be placed under this one. It looks as if the predominant, though never completely triumphant, urge is to accept and adapt to the usages of a socially arranged world. Then, gradually, more space is required for the description of cognitive and manipulatory developments in which an established order is being challenged by a more individualized organism interested in specific events. Then, towards the end, so far as this book goes, the social and the general seem to be beginning to come back again; not in the same way, though, but rather as if the main preoccupation now is to find how one increasingly individual organization of needs and tendencies can maintain and extend itself together with others. It is perhaps a pity that Prof. Valentine rings down the curtain just when he does. Maybe if he had carried on the play for a few more years the full shape and significance of his characters would have become easier to see. He has done enough, however, in all conscience, to merit prolonged and, I hope and believe, lasting applause.

PUBLICATIONS RECENTLY RECEIVED

Modern Psychology. By CHARLES S. MYERS, FRANK N. FREEMAN and MORRIS S. VITELES. Philadelphia: University of Pennsylvania Press (London: Humphrey Milford). 1941. Pp. 48. 3*s.* 6*d.* net.

Although the title of this publication is hardly justified by its contents, the book is of special interest to readers in this country because nearly half of it is taken up by a lecture by Dr C. S. Myers, entitled Aspects of Modern Psychology. In this lecture, Dr Myers passes in review the development of psychology from the time of Fechner to the present day. Nobody could do this better, for wide and accurate knowledge is combined with an acute yet tolerant criticism. Dr Myers moves easily and surely in all the fields of psychology which he discusses, experimental, abnormal, descriptive, or that very crowded and tangled contemporary world of 'tests'. The lecture is one which ought to be brought to the notice of intelligent students everywhere. It will tell them about the many views which have been and are current in psychology without leaving the impression of muddle and hopeless conflict which many such reviews produce.

Prof. Freeman's lecture on Educational Adjustments is less happy. He discusses Mental Growth and Development, Learning and Individual Differences, but all in a rather perfunctory way after the style of one who has to produce a lecture, but does not very much want to do so.

Prof. Viteles, taking practically the same amount of space, contrives to impart a great deal more information and to make some valuable suggestions as to the lines along which the Application of Psychology to Industry and Business is likely to travel. His chosen topics are: accidents and their prevention, and sources of satisfaction and dissatisfaction in relation to work, with particular reference to effective incentives. The discussion is concise and comprehensive, exceedingly lucid and well arranged, and Prof. Viteles writes as if he himself believes that his problems are of great importance in contemporary life, as they undoubtedly are.

All three lectures were contributions to the University of Pennsylvania Bicentennial Conference.

The Pub and the People. By 'MASS-OBSERVATION'. London: Victor Gollancz, Ltd. 1943. Pp. 350. 16*s.* net.

It is essential to remember in reading this book that it belongs to the earlier stages of 'Mass-Observation's' career. The book was in proof in 1939, and therefore the collection of data, which seems to have taken about two years, must have occurred at a considerably earlier period. Thus it is not unlikely that the authors felt they were groping for a path in semi-darkness among a luxuriant proliferation of miscellaneous data; at least, this is somewhat the impression received by the reader. It is clear that no team of observers could observe *all* the pub behaviour in Worktown *all* the time, even over a short period. They had to sample the behaviour to be observed; and they had then to present their sample observations to the reader in the form of a coherent pattern unless they were content to leave them in a form as meaningless as that of the kaleidoscope pattern. And herein lay the difficulty. The writer who arranges a coherent pattern is bound to have certain preconceived notions, of which he is more or less clearly aware, which direct and to some extent determine not only the nature of the pattern but also the selection of data which form it. The authors of *The Pub and the People* were apparently uncertain whether (1) to leave their sample observations in kaleidoscopic form, or (2) to use them to demonstrate the thesis that the pub forms an essential feature in the spontaneous social groupings and relationships of the working classes. It is because they did not choose clearly either the one course or the other that the reader is left with a feeling of confusion; ready to entertain the aforesaid thesis, but uncertain how far the plethora of observations, conversations, reflexions, etc., really support this thesis. There are further hints that the pubs and the rituals of drinking are complementary to, or have replaced, the sacrificial rites of the Church. But these are very vague, and, in the absence of some observed basis in unconscious psychological tendencies, appear meaningless.

The above dilemma has been described at some length, because it is one which must confront any social psychologist or anthropologist. 'Mass-Observation' was inclined originally to gloss it over; to confine itself to the kaleidoscopic method, or to assume that its observations must be entirely objective, because the observers were unaware of possessing any obvious preconceived theory on the subject. But in what has been published since the war began, a clearer thread of argument has emerged; and the mass of data have fallen into place in demonstrating this argument.

Nevertheless, this book, though it would have been improved by curtailment, does present evidence in support of the preservation of the 'local' as an institution essential in working-class life—one where the working man can relax, meet his fellows and enter into a satisfying equalitarian relationship with them. This is contrasted with the non-social, if not anti-social, influence of the cinema and the football pools.

M. D. V.

Psychology for Individual Education. By Lois Barclay Murphy, Eugene Lerner, Jane Judge and Madeleine Grant. Edited by Esther Raushenbush. Sarah Lawrence College Publications, No. 2. New York: Columbia University Press (London: Humphrey Milford). 1942. Pp. xii + 306. 18*s.* 6*d.* net.

It is the belief of the Director and teachers of the Sarah Lawrence College, New York, that the education of students, especially at the beginning of their college careers, should be designed in such a manner as to help them develop in accordance with their individual needs and aptitudes, "to achieve as high a degree of emotional and intellectual maturity as possible". Each student is therefore expected to take an "exploratory course" during her first year to enable her to find out for herself what is her natural bent—and also to give the teachers an opportunity to find out what kind of a person she is, and what type of education would suit her best. On the basis of school reports, interviews and consultations with advisers, it is decided which of the possible exploratory courses is likely to be most suitable for her. Many students are benefited by courses in elementary psychology, particularly the psychology of personality, aimed to help them to "find out about people" or to understand themselves, their friends and families. Others may prefer a more 'scientific' approach to the study of psychology, or may like to tackle personality and behaviour problems less directly; for them a biological or a sociological approach is often valuable.

This book describes in detail four such psychological courses and the reactions of certain students to them. There is little doubt that the courses indicate a most interesting and novel method of tackling individual education and development in relation to society. But they would be rather difficult to fit into the educational curriculum in this country, where the age of entry to college is usually higher, for they appear to fall midway between a sixth-form school course and a first-year university course. Nevertheless, any educationist not hag-ridden by the examination spectre should find a great deal that is useful and suggestive in this method of suiting the education to the individual, rather than the individual to the education.

Young Children in War-Time. By Dorothy Burlingham and Anna Freud. London: George Allen and Unwin, Ltd. 1942. Pp. 81. 1*s.* 6*d.* net.

This pamphlet begins with a brief account of the establishment of the residential war nursery known as the Hampstead Nursery. It was started under the auspices of the American Foster Parents' Plan for War Children, which originally provided 85 % of the budget; but this support was greatly reduced when America entered the war, and the nursery is now dependent to a considerable extent on voluntary contributions from this country. It takes in 50 children, mainly aged 5 and under, whose homes have been broken up by the blitz, through parents being in the Services or out of work, etc., or who have been unsuccessfully evacuated and returned to London. There is a day nursery for these and 40 other non-resident children; and a country residential nursery, to which some 30 children have been evacuated. A trained medical, psychological and educational staff is in attendance at the nurseries.

The pamphlet then gives a non-technical account of the psychological observations made (mainly it appears under the guidance of Dr Anna Freud) as to the effects on these children of bombing and evacuation. Previous observations are confirmed that the child of under 5 years was comparatively little affected by bombing, provided that he himself was not physically injured, that he was in his mother's care, and that she did not show undue signs of panic. The effects of separation from parents through evacuation seem to have been much more far-reaching, although the child might show little obvious signs of them because at that age he is unable to talk about his unhappiness and his sense of desertion and rejection. He may in time transfer his affections to foster parents, teachers or nurses, but with the further unfortunate consequence that he no longer wants his own parents. It is therefore recommended that if possible the child should go first to a London residential nursery, where he can still see his parents from time to time, before being sent away to the country. Thus the separation from the mother is gradual, the shock to the child is less and hence also his feelings of rejection, and his own reactions of rejection of his parents.

The authors of this pamphlet put forward a most interesting thesis as to these psychological effects on the child. But it cannot be taken as proven, since the cases are comparatively few in number (about 100), and the actual frequency of the examples cited is not shown. Moreover, it sometimes appears that observation and interpretation of behaviour are made to fit theory, rather than vice versa. But there is no doubt that the Hampstead Nursery is doing most valuable work, both in assisting these children to recover from the effects of bombing and evacuation, and also in studying the underlying psychological basis of these effects.

Preliminary Report on Children's Reactions to the War. By J. Louise Despert. From the New York Hospital and the Department of Psychiatry, Cornell University Medical College, New York. 1942. Pp. 92+10.

The first part of this monograph consists of a detailed account and discussion, with full references in the Bibliography at the end of the monograph, of previously published literature on children's reactions to the war, bombing, evacuation, etc. The list of references is extremely useful for those interested in this subject, but the abstracts themselves are not always very clear and helpful in giving the essential results; nor are they always entirely accurate.

The second part of the monograph describes a study of the general behaviour, play and conversation relating to the war of children at an American nursery school; the psychologist's own observations were supplemented by reports from parents. It was found that anxiety relating to the war situation occurred only in children who had previously presented anxiety problems—in general, those who were insecure in their relationship to one or both parents. This is rather to be expected, since these children had not been subjected to any direct war strain. The younger children were not affected because they had insufficient intellectual insight into war conditions.

Child Art. By Wilhelm Viola. London: University of London Press Ltd. 1942. Pp. 191 +Plates XII. 15*s.* net.

This book, written by a disciple of the Austrian art teacher, Cizek, is intended to prove effectively to anyone who still doubts the fact, that children should be allowed to express themselves freely in their drawing, painting and modelling, and should not be urged or compelled to copy accurately from models. Whether the children's productions should indeed be called 'art' is another matter. But there is no doubt how much the child is benefited by the opportunity and encouragement for this free expression; and children, their parents and teachers, owe a debt of gratitude to Cizek for originating and propagating his methods in the face of great discouragement.

Cizek's methods are described and discussed in detail, and there is a long series of answers to questions about these methods and their effectiveness, and of conversations between Cizek and his child pupils. The book is somewhat long-winded for the ordinary reader, but should be of much value to all who are professionally interested.

Self-Analysis. By Karen Horney. London: Kegan Paul, Trench, Trubner and Co., Ltd. 1942. Pp. 309. 10*s.* 6*d.* net.

The title of this book is a little misleading, and may discourage people from studying it, as they might do with much interest and profit. The book consists, first, of an excellently clear and interesting analysis of the process of psychoanalysis (not of the strictly Freudian variety), with a discussion of the main neurotic trends and the manner in which they may be uncovered and dispelled. And this is followed by a discussion of the possibilities of the patient's continuing the analysis by himself without the analyst's assistance or with only occasional consultations. This is illustrated by the detailed consideration of a single case in which excellent results were obtained, and by brief descriptions of other cases in which it was successful up to a point. It does not appear that Dr Horney has any evidence for the possibility of successful self-analysis, unless of a very superficial kind, which has not been preceded or supplemented by the treatment of an analyst. Indeed, she herself advances arguments which make this possibility appear unlikely. Even a highly intelligent and energetic individual, possessed by the strongest motive or desire to find out the truth about himself, would be unlikely to be able to work out the full emotional implications of his mental conflicts and neurotic trends without assistance from the encouragement and the insight of the analyst—and still more, without the opportunity to focus these conflicts on the person of the analyst and 'live through', in the transference, their full emotional significance.

Nevertheless, the book can be strongly recommended as a most readable study of psychoanalysis, both from the practical point of view (i.e. as to the desirability and efficacy of analysis), and also from the theoretical aspect, the light which it throws upon psychological processes in general.

Man the Master. By Gerald Heard. London: Faber and Faber, Ltd. 1942. Pp. 249. 10*s.* 6*d.* net.

This book consists of a further detailed discussion of Gerald Heard's thesis that the only issue from our present discontents will be through the guidance of the 'seer', the individual whose insight into the unconscious minds of himself and his fellows has been developed and enlarged to its fullest extent.